University of Groningen

Understanding and managing interruptions

Katidioti, Ioanna

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Katidioti, I. (2016). *Understanding and managing interruptions: How to avoid watching cat videos all day long*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 23-05-2025

Understanding and Managing Interruptions

How to avoid watching cat videos all day long

Ioanna Katidioti

Printed by, NetzoDruk
NetzoDruk, the Netherlands

ISBN printed version: 978-90-367-8978-3
ISBN digital version: 978-90-367-8977-6

This investigation was partially supported by a grant from the European Research Council (ERC) awarded to Niels Taatgen (grant no. 283597).

The cover was designed by Myrto Stathatou
I thank Simon's Cat© for allowing me to use their art for my thesis.

rijksuniversiteit
groningen

Understanding and Managing Interruptions

How to avoid watching cat videos all day long

PhD Thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 24 June 2016 at 12.45 hours

by

Ioanna Katidioti

born on 7 May 1985
in Marousi, Greece

Supervisor
Prof. N.A. Taatgen

Co-supervisors
Dr. J.P. Borst
Dr. M.K. van Vugt

Assessment committee
Prof. G.J. Mark
Prof. E.M. Steg
Prof. C.D. Wickens

Table of contents

Part 1

Part 2

Part 3

CHAPTER 1

INTRODUCTION

INTRODUCTION

This chapter has 3258 words. How long does it take to write and revise a text of 3247 words? One working day, eight whole hours of typing and reading, seems more than enough for this amount of text. However, I think no one will find it strange when I admit it took me much more than one day to do it.

When trying to work at my office, it feels as more time is spent talking to colleagues, checking my emails or just browsing aimlessly online than actually writing. When working from home, I avoid colleagues and can turn off email and social media notifications, but I still have to deal with my aimless Internet browsing, my cat wanting to be petted, or my book frequently calling me to read it.

For many people, it may seem impossible to focus on one task for long without being interrupted, either by an external or an internal source. External interruptions are sometimes unavoidable, for example answering the phone might be a required part of your job, even if the call occurs in the middle of writing something. External interruptions can occur at the most inconvenient moments, but there are ways to avoid or at least minimize them, for example by closing your office door or turning off your cellphone. Self-interruptions are harder to manage. You know you should be focusing on your work, but for some reason you keep checking the news website every 10 minutes. To keep self-interruptions under control, more and more people use special apps that block their access to the Internet or they time themselves (e.g., in the pomodoro method you work for a specific amount of time before taking a break) in order to reduce self-interrupting.

It is clear to most of us that interruptions affect our performance in a negative way. We make more errors, take more time to complete our main task or we even postpone its completion. When asked, most people say that interruptions should be eliminated completely in order to achieve optimal performance. They may even consider locking themselves away in a place without Internet connection until the task is finished. In contrast, in this thesis I will accept interruptions as an unavoidable part of everyday life. Even if I lock myself in a room without Internet connection in order to write this thesis introduction, I will still stare at the wall from time to time wondering what to have for lunch or what I should wear for my thesis defense 4 to 5 months from now. Interruptions will keep happening no matter how much we try to avoid them. Therefore, in this thesis I will try to gain a better understanding of the causes behind interruptions and the best way to manage them.

In the first part of this thesis I will try to discover more about self-interruptions. By means of behavioral and eye-tracking experiments, I will examine how cognitive resource availability (cognitive resources, such as vision, are said to be "available" when they are not used by a task) can affect distractibility and interruptibility. I will answer questions such as: What happens to people's rational self-interrupting behavior when they are faced with a browser delay at a moment that it would not be rational to interrupt their main task? Will they self-interrupt or not? The next step will be to move deeper into studying the effects of cognitive resource availability on distractibility. As their main task gets harder, will people be more or less distracted by a cat video that keeps playing in the periphery of their visual field? Does that answer depend on what kind of task they perform: one that needs more visual resources as it becomes more difficult versus a task that requires more cognitive resources as it becomes more difficult?

The second part of this thesis will compare self-interruptions to external interruptions, by answering questions such as: What happens before a self-interruption? Is there a difference with what happens before an external interruption? Which kind of interruption is more disruptive? Finally, in the third part of this thesis, I will present an interruption management system that makes use of our acquired knowledge by trying to interrupt people during moments that are least disruptive, using pupil dilation as a physiological indicator of those moments.

In this Introduction section, I will first present some background on interruptions, the factors that influence their disruptiveness, and existing theories on interruptions. I will then mention how cognitive resources affect interruptions and how there is a need for experimental studies that directly compare self-interruptions and external interruptions. Finally I will present pupillometry as a research method in cognitive science, a method I used in several of the studies reported in this thesis.

BACKGROUND

Definition of interruptions

Interruptions can be considered as a form of sequential multitasking (see Salvucci & Taatgen, 2011), since the person that is interrupted has to deal with more than one tasks. Figure 1 shows the timeline of interruptions (adapted from Trafton, Altmann, Brock, & Mintz, 2003). First, someone performs a main task, such as writing a thesis introduction.

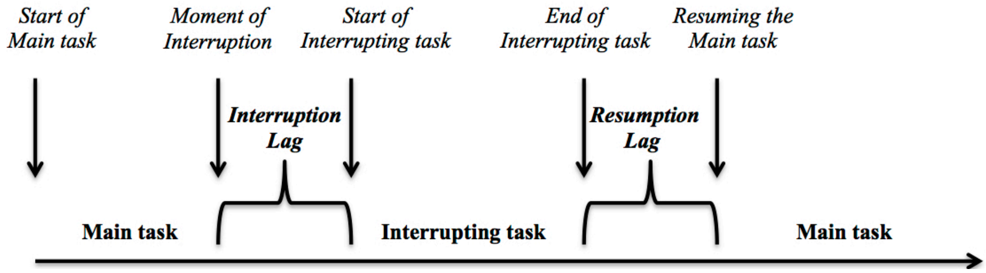


Figure 1.1. *Interruption timeline, based on Trafton et al. (2003)*

An interruption occurs, for example the sound of an incoming email. The time between hearing the sound and opening the email is called the interruption lag. Next, the interrupting task (reading the email) is performed. The time between having read the email and resuming the writing of the introduction is called the resumption lag. Finally the main task is continued.

Trafton et al. (2003) assume on the basis of this timeline that the total cost of interruption is the sum of the interruption lag and the resumption lag (naturally, this is in addition to the time lost to the interruption itself, but that time might be spend effectively on a secondary task). However, there are more factors that could contribute to the disruptiveness of the interruption, such as loss of efficiency after the main task is resumed. In this thesis I will not focus on the resumption lag as a measure of disruptiveness, but I will look at performance on the whole task. Furthermore, I will investigate whether this timeline is also valid for self-interruptions.

Frequency and effects of interruptions

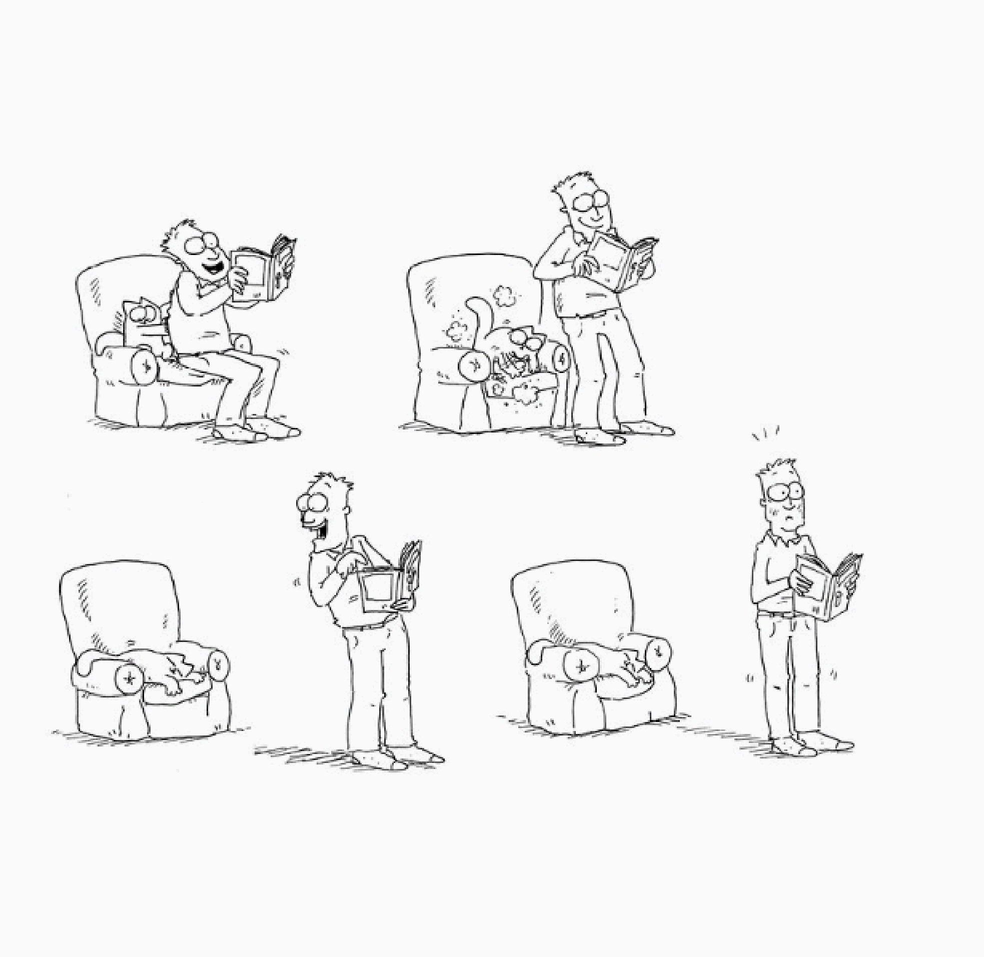
There are more than a few studies that confirm what we all suspect: interruptions happen constantly, with self-interruptions being as frequent as external interruptions. For example, a large observational study revealed that information workers are interrupted every 3 minutes and about half of these interruptions are self-interruptions (Gonzalez & Mark, 2004). Interruptions interfere with important aspects of everyday life, such as working (e.g. Mark, Gonzalez & Harris, 2005) and studying (e.g. Rosen, Carrier, & Cheever, 2013). The main negative effects of interruptions are that the main tasks take more time to be completed and more errors are made (e.g., Monk, Trafton, & Boehm-Davis, 2008; Brumby, Cox, Black & Gould, 2013).

However, the disruptive effect of interruptions varies. There are a number of factors that can make an interruption more or less disruptive. First, long interruptions are more disruptive than short ones (e.g. Gillie & Broadbent, 1989; Monk et al., 2008). Second, complex interrupting tasks are more disruptive than simple ones (e.g. Cades, Boehm-Davis, Trafton & Monk, 2007; Monk et al. 2008). Third, an interrupting task relevant to the main task is less disruptive than an unrelated one (e.g. Czerwinsky, Cutrell, & Horvitz, 2000; Gould, Brumby, & Cox, 2013). Fourth, an alert before the interruption minimizes the resumption lag (e.g. Hodgetts & Jones, 2003, Monk et al. 2008). Finally, the moment that the interruption occurs within the main task affects the level of disruption. An interruption that occurs when the cognitive workload of the task is high (e.g. while in the middle of typing a word) is more disruptive than an interruption during a low-workload moment, such as after finishing a paragraph (e.g. Iqbal & Bailey, 2005; Monk, Boehm-Davis & Trafton, 2004).

Despite all the literature mentioned above, much remains to be discovered about interruptions. One area that has seen little research are self-interruptions: there are very few experiments on self-interruptions and the underlying causes of our constant need to interrupt our work to engage in unnecessary tasks are still unknown. Furthermore, there is the need for a cognitive theory that can provide a satisfactory explanation for all the experimental results mentioned above.

Interruption theories

One of the main theories on interruptions is Memory for Goals (Altman and Trafton, 2002). According to this theory, each task has a goal and an activation level. When the main task is interrupted, its goal is stored in memory and starts to decay, while the interrupting task's goal is activated. Returning to the main task entails resuming its goal. Longer interruptions result in a greater decay of the main task goal and consequently it takes more time to resume the main task after a longer interruption. Although Memory for Goals takes into consideration the effects of the length of the interruption, it does not account for the other factors that affect the disruptive effect of interruptions such as resource availability.



This strip by Simon's Cat© was a low-workload moment interruption

Borst, Taatgen and van Rijn (2015) extended Memory for Goals theory to Memory for Problem States and shifted the focus from goals to problem states (problem state contains the information necessary for the completion of a task). They propose that the main and the interrupting task could have an associated problem state, in which case the interruption is handled as in Memory for Goals: the main task's problem state is stored and starts decaying until the interrupting task is finished. However, if the main or the interrupting task does not require a problem state, the length of the interruption will not affect the resumption lag. With this extension, the factors of interrupting task complexity and moment of the interruption are also taken into account.

Interruption theories are improving, but they still cannot account for all phenomena. One challenge is self-interruptions, probably because they are difficult to evoke in controlled experiments. The results of this thesis will extend our knowledge and theories of interruptions.

Cognitive resource availability and interruptions

In the first part of this thesis I will focus on the effect of cognitive resource availability on self-interruption. Cognition can be divided into separate cognitive resources, such as vision, working memory, motor functions etc. and different tasks make use of different cognitive resources. For example, watching a silent video uses visual resources, while writing the introduction of your thesis uses vision, declarative memory, language processing and motor resources.

It is well known that cognitive resource availability affects multitasking behavior. According to Wickens' Multiple Resources Theory (2002) there is greater interference when one is trying to combine two tasks that share a cognitive resource than when trying to combine two tasks that do not share a resource. People cannot look simultaneously at two things; therefore combining two visual tasks is hard, whereas combining two tasks that do not share a resource (e.g. driving while listening to music) is easy (this is also part of Threaded Cognition theory; Salvucci & Taatgen, 2011).

One of the principles mentioned in Salvucci and Taatgen's Threaded Cognition theory on multitasking (2008; 2011) is the resource usage principle, according to which resources are being used in a greedy, polite manner. That means that if two tasks are competing for the use of a resource, once this resource becomes available from the one task it will be released ("polite") and the other task will use it ("greedy").

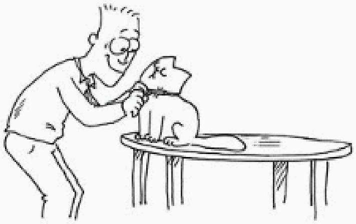
For example, when the silent video you are watching in your browser ends, your visual resources will be occupied by the advertisements on your screen or the suggestions for similar videos. Using the theories of Threaded Cognition (Salvucci & Taatgen, 2011) and Multiple Resources (Wickens, 2002) on self-interruptions, the first part of this thesis turns the logic around, and tries to answer the question: when a cognitive resource becomes available, will that lead to self-interruption?

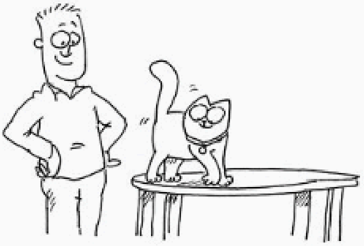
There are experiments that demonstrate the ability of people to self-interrupt rationally (e.g. Salvucci & Bogunovich, 2010). When given the freedom to self-interrupt, people usually chose optimal moments and minimized the negative effects of self-interruptions. However, I hypothesize that once the resources needed by the interruption are available, people will self-interrupt at the wrong moments. I will test this cognitive resources theory by using a visual distraction during two different tasks: one that releases more visual resources as it becomes more difficult and one that uses more visual resources as it becomes more difficult. On the basis of my hypothesis, the prediction is that participants will be more distracted by the visual distractor as the visual resources become more released, even if the main task becomes more difficult.

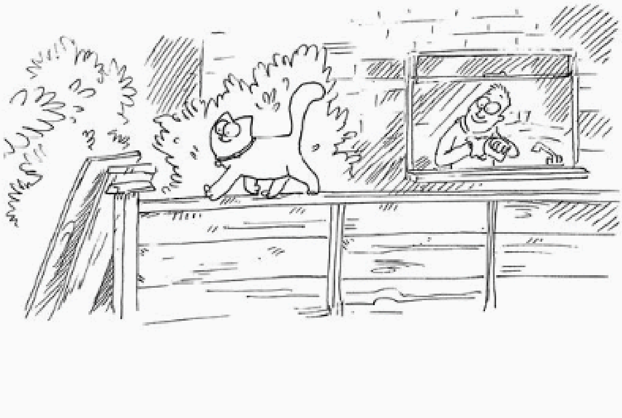
External vs. self-interruptions

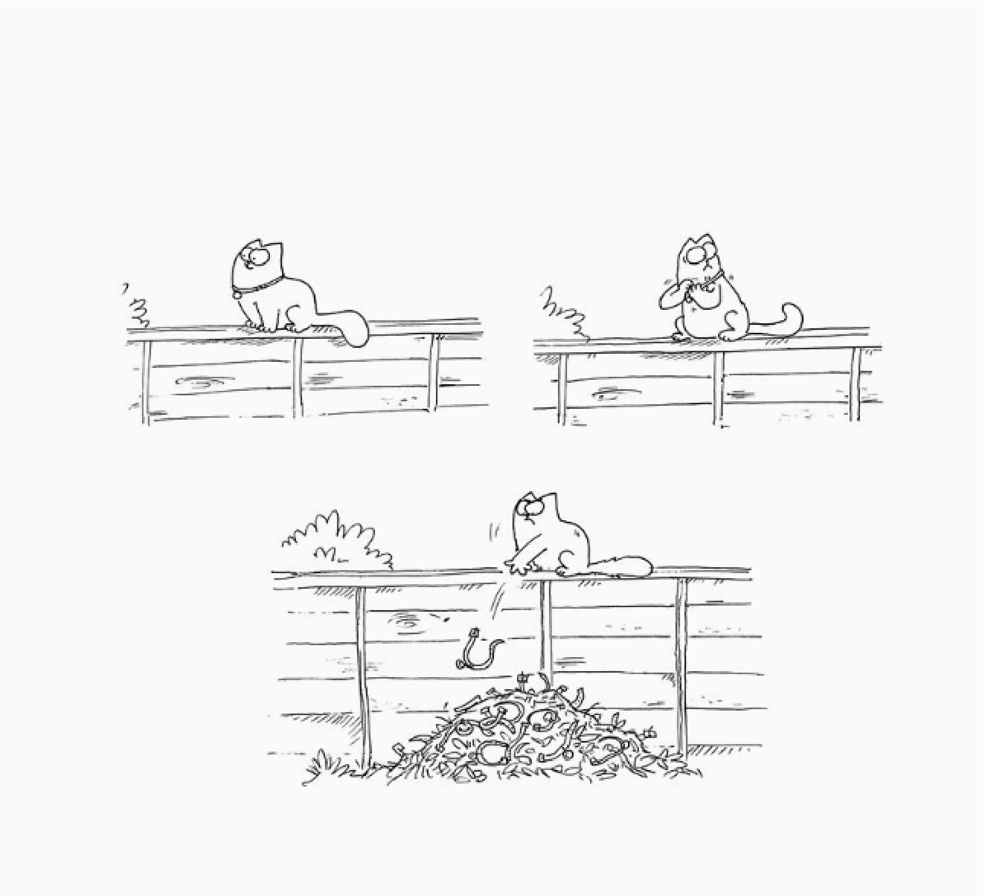
The second part of this thesis will focus on comparing self-interruptions and external interruptions. Although there are many experimental studies showing the negative effects of external interruptions (e.g. Altman & Trafton, 2002; Monk et al., 2008), there are only a few investigating self-interruptions. Although observational studies reveal that self-interruptions are as common and as disruptive as external interruptions (e.g. Dabbish, Mark & Gonzalez, 2011), it is not simple to create an experimental setup for studying them. However, one consistent result of the few existing studies that give participants the freedom to choose their interruption moments (e.g., Payne, Duggan & Neth, 2007) is that people usually behave rational when self-interrupting: People will seldom choose to check their emails whilst in the middle of typing a word. They will wait for the end of a paragraph or a chapter in order to do that and that way they will resume their main task more easily.

There are even fewer studies that compare self-interruptions with external interruptions and even when they do, the design is usually flawed. For example, McFarlane (2002) compared four different kinds of interruptions, three external (immediate, scheduled and using an









This strip by Simon's Cat© was a high-workload moment interruption.

Was it more disruptive than the previous one?

interruption management system) and one negotiated interruption. However, his negotiated interruption was not exactly a self-interruption, since participants were externally interrupted by the interrupting task, but could choose when to act on that interruption. Panepinto (2010) gave participants the freedom to self-interrupt during a Sudoku game, but the external interruptions in her study occurred at random moments. Since participants probably chose the optimal moments to self-interrupt[1] the comparison automatically favors the self-interruptions. However, Panepinto did not find a difference in performance between the two kinds of interruptions. Assuming that forced interruptions occurred on average at more disruptive moments, this suggests that forced interruptions might in fact be preferable over self-interruptions.

In order to compare external and self-interruptions, both kinds of interruptions should occur at similar moments within the task. For that reason, I created a situation in which both external and self-interruptions occurred at the same moments of the main task in Chapter 4 and Chapter 5 of this thesis. I hypothesize that if both interruptions happen on rational moments (when the workload is low), external interruptions will be less disruptive, since they do not require the person to make a decision on when to interrupt themselves.

METHODS

Pupil Dilation

In this thesis I will use changes in pupil dilation as a method to study the effects of interruptions. Pupillometry has been used widely for decades in cognitive science (e.g. Beatty & Lucero-Wagoner, 2000). It is common knowledge that the pupil reacts to light changes, however, many studies show that it reacts to non-visual stimuli as well, such as emotions and cognitive processes. Pupillometry has been used to study Stroop effects, task complexity, workload changes and more (e.g. Laeng, Sirois & Gredeback, 2012). Pupil dilation also reflects some more unexpected cognitive functions, such as uncertainty when gambling (Satterthwaite, Green, Myerson, Parker, Ramaratnam & Buckner, 2007) or suppression of the urge to press a button (Chiew & Braver, 2013).

In this thesis, pupil dilation will be used mainly as a measure

---

[1] There is no such mention in Panepinto's paper, but given other studies (e.g., Salvucci & Bogunovich, 2010), it seems natural to assume that.

of cognitive workload[2]. Since the 1960s, there have been many studies showing that pupil dilation increases as the mental workload increases (Kahneman & Beatty, 1966). For example, many studies showed that pupil dilation increases as the number of elements that need to be retained in the working memory increases (e.g. Peavler, 1974; although of course there is a limit to this phenomenon, as Granholm, Asarnow, Sarkin & Dykes, 1996 suggest). There are also different kinds of increases in mental workload that create increases in pupil dilation: mathematical equations (e.g. Hess & Polt, 1964) or language tasks (e.g. Schluroff, 1982) create a bigger increase in pupil dilation as they become more difficult.

Interruption studies have made use of this relation between pupil dilation and cognitive workload. For instance, Iqbal, Adamczyck, Zheng & Bailey (2005) used pupil dilation to identify low-workload moments of a document editing and a route-planning task. In a follow-up study Iqbal & Bailey (2005) used these defined low-workload moments to schedule interruptions and discovered that (as expected) interruptions at low-workload moments are less disruptive than interruptions at high-workload moments. In the interruption management system I present in Chapter 6, I use the fact that pupil dilation decreases at low-workload moments to create less disruptive external interruptions by automatically interrupting users at low-workload moments. This goes beyond previous systems in that it does not require a task analysis.

Furthermore, in Chapter 4 and Chapter 5 I investigate how pupil dilation reacts before a self-interruption and an external interruption, and what that tells us about the difference between external and self-interruptions. In these chapters I use pupil dilation changes mainly to compare the timeline of a self-interruption with that of an external interruption. Pupillometry has been used in many studies in order to identify the different parts of a task (e.g. Hess & Polt, 1964, Iqbal et al., 2005) and is a useful method for finding differences between cognitive reactions.

---

[2] To be more precise, what studies refer to as “cognitive/mental workload” would be better described as “processes related to cognitive control”. For consistency with past research, we will keep calling it “cognitive/mental workload”.

DISSERTATION OVERVIEW

The area of focus of this dissertation is self-interruptions, and it seeks to answer the following questions:

- How does cognitive resource availability affect interruptibility?
- What happens before a self-interruption?
- Are external interruptions more or less disruptive than self-interruptions?
- How can we minimize the negative effects of interruptions?

Part 1 of this thesis focuses on how cognitive resource availability affects interruptibility and consists of Chapter 2 and Chapter 3. In Chapter 2 I present an experiment which questions people's rationality in self-interrupting, by enhancing cognitive resource availability on moments that it is not rational to self-interrupt. In Chapter 3 I discuss the effects of cognitive resource availability on distractibility by presenting an experimental study that compares two tasks that use different cognitive resources.

Part 2 of this thesis focuses on comparing external and self-interruptions and consists of Chapter 4 and Chapter 5. Chapter 4 focuses on the differences between the effects of self-interruption and external interruption on pupil dilation. Chapter 5 again compares self-interruptions and external interruptions in order to replicate the results of Chapter 4 on a different task and confirm which kind of interruption is more disruptive.

Part 3 of this thesis consists of Chapter 6 and describes an interruption management system that tries to interrupt people on optimal moments based on changes in pupil dilation to cognitive workload. Finally, Chapter 7 offers an overview of the results of this thesis and the insights they offer on interruptions.

PART 1

The effects of cognitive resource availability on self-interruptions

CHAPTER 2

Choice in Multitasking:
How delays in the primary task turn a rational into an irrational multitasker.

This chapter was previously published as:
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

ABSTRACT

Objective: To establish the nature of choice in cognitive multitasking.

Background: Laboratory studies of multitasking suggest people are rational in their switch choices regarding multitasking, while observational studies suggest they are not. Threaded cognition theory predicts that switching is opportunistic and depends on availability of cognitive resources.

Method: 21 participants answered emails by looking up information (similar to customer-service employees) while being interrupted by chat messages. They were free to choose when to switch to the chat message. We analyzed the switching behavior and the time they needed to complete the primary mail-task.

Results: When participants are faced with a delay in the email task, they switch more often to the chat task at high-workload points. Choosing to switch to the secondary task instead of waiting makes them slower. It also makes them forget the information in the e-mail task half of the time, which slows them down even more.

Conclusion: When many cognitive resources are available, the probability of switching from one task to another is high. This does not necessarily lead to optimal switching behavior.

Application: Potential applications of this research include the minimization of delays in task design and the inability or discouragement of switching on high-workload moments.

INTRODUCTION

Multitasking has become more important in modern office work and life in general. People multitask in their cars, on the street and while working on their computers. A key property of many multitasking situations is that people *choose* to multitask, deciding themselves to carry out tasks at the same time, or to switch from one task to another without direct external reason. The goal of this study is to investigate the process of deciding to switch from one task to another.

Observational studies have demonstrated that modern office workers switch between tasks often, on average every 3 minutes in a study by Gonzalez and Mark (2004), and typically require substantial amounts of time to return to a main task after they have been interrupted (Mark, Gonzalez & Harris, 2005). Estimates have been made that 28% of a knowledge worker's day is spent on interruptions (Spira & Feintuch, 2005). Although these studies suggest multitasking leads to a loss of productivity, they cannot establish this with certainty.

The costs of multitasking have been demonstrated in several laboratory studies, and are higher when mental workload is higher at the moment of interruption (Adamczyk & Bailey, 2004; Czerwinski, Cutrell & Horvitz, 2000). The costs of interruptions can be explained by memory decay: the mental representation of an interrupted task becomes less available over time, leading to additional time requirements to either recall or reconstruct the interrupted task (Altmann & Trafton, 2007).

A limitation of almost all laboratory studies is that the interruption is forced. Gonzalez and Mark (2004), however, found that half of the interruptions observed in real office situations are self-initiated, which means that some aspect of the current or new task prompted people to switch to a different task. Even though there are some studies that show that people tend to stabilize their task before they switch (e.g., Iqbal & Horvitz, 2007), it is not yet clear what the extent of people's rationality is regarding self-interruption.

There have been some experiments that examine what happens when people have some freedom in choosing when to switch. Sellen, Kurtenbach & Buxton (1990) showed that people tend to delay switching to another task until they complete a subtask. In a study by McFarlane (1999) participants performed better in a condition in which they were allowed to choose the moment of interruption compared to conditions where they were forced to switch immediately.

The experiment we will discuss here is an extension of an experiment by Salvucci and Bogunovich (2010), where participants had the freedom to choose when to interrupt themselves. Participants had to answer emails as a primary task and respond to chat messages as a secondary task. They were free to switch to chat messages whenever they wanted. Salvucci and Bogunovich's results showed that people made rational choices: 94% of the switches to the chat-task were made at low-workload points in the mail-task. We define rational choice as in Anderson's (1990) principle of rationality, where the cognitive system tries to optimize the behavior to fit the demands of the environment, taking into account the limitations of the cognitive system.

To summarize, the two sources of information with respect to choice in multitasking contradict each other: observation studies suggest people make poor choices in multitasking, while laboratory studies suggest choices are rational.

To explain choice in switching, we will consider two possible accounts. The first is based on *utility*, and predicts people switch because it will lead to the best payoff in the least amount of time (we can call this global rationality). The second is based on *resource availability*, and predicts that people switch when they have many cognitive resources available (we can call this local rationality).

One prominent utility account is the soft constraints hypothesis of Gray and colleagues (Fu & Gray, 2006; Gray, Sims, Fu & Schoelles, 2006). According to this theory, people optimize their choices in switching within a single task to maximize utility. Although not designed for multitasking, it fits the existing experimental results in multitasking very well. However, there is a conceptual problem in applying soft constraints to multitasking. The assumption of utility-based accounts is that utility is a property of the knowledge for that task, and that whenever there is a choice between two knowledge elements, the element that leads to the highest payoff for that task in the least amount of time is chosen. This means, however, that switching to another task is never attractive from a utility perspective because it only delays achieving the goal. Successful use of a utility strategy would require maintenance of a global utility in addition to utilities for the separate tasks, which may not be a tractable solution.

A resource availability account is provided by the threaded cognition theory (Salvucci & Taatgen, 2008, 2011). Threaded cognition is designed specifically to model multitasking. It assumes cognition can be subdivided into several separate cognitive resources (vision, motor, declarative memory, working memory, etc.). Particular tasks typically only use a subset of the resources and often only for certain periods of time, which means multiple tasks can be carried out without interference as long as they do not require the same resource at the same moment in time. The decision process is very simple: if a task needs a particular resource, and that resource is currently not in use by another task, it can use it.

The consequence of threaded cognition's decision process is that choice in multitasking depends on the availability of cognitive resources. If all cognitive resources are engaged with a task, the probability of switching is low, but if resources are not in use, there is a tendency to add or switch to a task that requires those unused resources. This behavior is locally rational, because it tries to use all resources as much as possible, but not necessarily rational with respect to global utility.

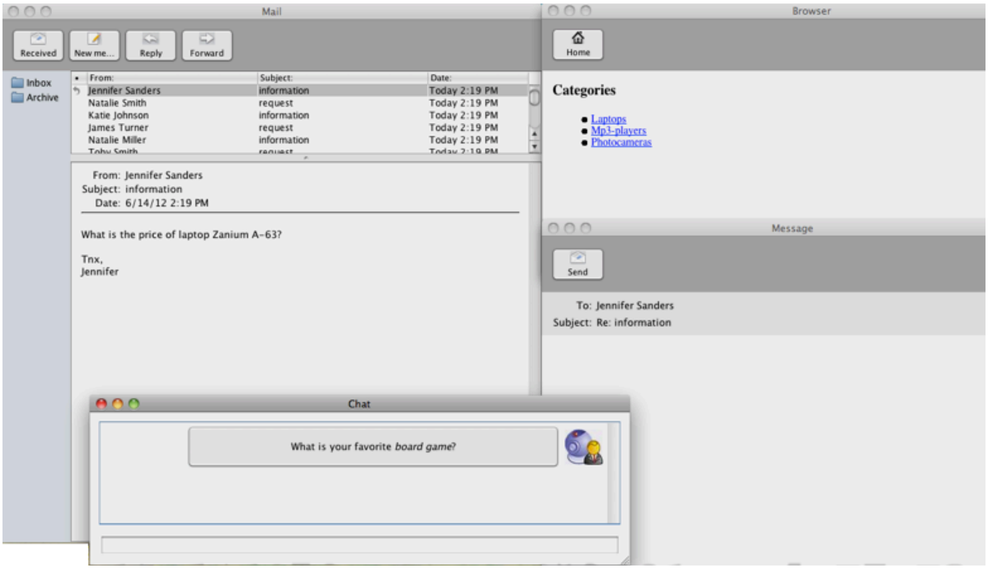
Threaded cognition supplies a different explanation for the rationality of choice in laboratory studies. In most multitasking experiments bad switch moments coincide with the primary task using more resources, making it less likely that a secondary task can intervene. This gives the impression people are globally rational about their multitasking. But according to threaded cognition, this rationality is based on the local optimization of resource use and therefore only local.

In the original Salvucci and Bogunovich study, global and local rationality also coincides, because high-workload periods were also the periods in which switching tasks was most detrimental to performance. The experiment we report here has added a condition to the experiment that breaks this symmetry: it introduces short periods in which the workload is very low, but switching in such periods leads to poorer performance. A utility account would predict that people still will not switch during these periods, possibly after a period of learning to learn the appropriate utilities, but a resource availability account predicts that people switch during such periods, even if this decreases their overall performance.

METHOD

Participants

Twenty-one participants (12 women), with a mean age of 23.5, were tested in all four counterbalanced conditions. There was a 5-10 minutes practice trial in the presence of the experimenter. The experiment lasted roughly an hour and 15 minutes.



From:	Subject:	Date:
Jennifer Sanders	information	Today 2:19 PM
Natalie Smith	request	Today 2:19 PM
Katie Johnson	information	Today 2:19 PM
James Turner	request	Today 2:19 PM
Natalie Miller	information	Today 2:19 PM

From: Jennifer Sanders
Subject: information
Date: 6/14/12 2:19 PM

What is the price of laptop Zanium A-63?

Tnx,
Jennifer

To: Jennifer Sanders
Subject: Re: information

What is your favorite *board game*?

Figure 2.1. *Screen shots of the mail, browser, chat and message windows. During the experiment the windows were overlapping, so that the participant had to actively switch to a window in order to see it.*

Experiment

The experiment consisted of a primary task and a secondary interrupting task. The primary task was a mail task that resembled the work of a customer-service employee and the secondary task was a chat message with personal questions.

In the primary task participants had to open emails asking them for the price of a specific product. For each email the participant had to read and memorize the type, brand and code of product (e.g., "Laptop Zanium A-63"). The names of the brands and codes were fictitious. After reading the email the participant had to switch to the web browser window, obscuring the email window. The initial page in this browser would show a list of product types (Laptop, Mp-3 Player or Photocamera), each of which would contain a link to a next page with a list of brands (e.g., "Zanium"), which were unique for each product type. Clicking on the appropriate brand name would bring the participant to a page with all the product codes for that brand (e.g "A-63"). The pages with brand names listed 3 items for the Mp3-Players and Photocameras categories and 4 for the Laptops category, and the pages with product code always listed 10 items. After clicking the third link with the product code, the message "searching for price" appeared for 3 seconds. After this delay, the participant could read the price of the product. They then had to return to the email window and press the “Reply” button. This resulted in the appearance of a new window (message window), where the participant could type the price and then send the message. The participant had to archive every answered message by dragging it to the “Archive” folder. That concluded the primary task after which the participant could move on to the next email.

At semi-random moments, a chat message would arrive, to which the participants could switch whenever they wanted. There was on average one chat prompt per mail task. Every time a new chat message arrived, there was a notification sound and the chat window (which would be in the background but with the edge of the window visible) turned yellow. Figure 2.1 shows the three programs (email window-message window, browser window and chat window). In the real experiment, the windows always overlapped so that only one was visible at the time. Although this aspect of the interface is rather artificial, it mimics a common problem in interfaces with overlapping windows when information from one window has to be used in another window from a different application, and the screen does not afford viewing both.

This experiment is identical to the original experiment by Salvucci and Bogunovich (2010) with three exceptions. In their experiment participants only had to click two links instead of three to reach the price information (they did not have the first step of product type). This third link increases the memory load to three items. Their experiment also did not include the delay before the price appeared. We added this delay in order to create a clear low-workload moment in the middle of the task, a moment where working memory does not contain either the product name or its price. The third modification is that we added another delay in the email program: whenever the participant would switch to the email window, it took three seconds to show the message in that window.

In addition to the basic task described above, we added two experimental manipulations to the design: the presence or absence of a 3 second delay after clicking the first two browser links (Delay/No Delay conditions) and the difficulty of the questions in the chat-task (Difficult/ Easy conditions). In the Delay condition, there would not only be a delay after clicking Link 3 in the browser, but after clicking any of the links. After the participant clicked the first (product type) and second (brand) link in the browser, a “loading” page would be shown for 3 seconds before proceeding to the requested page that the link referred to.

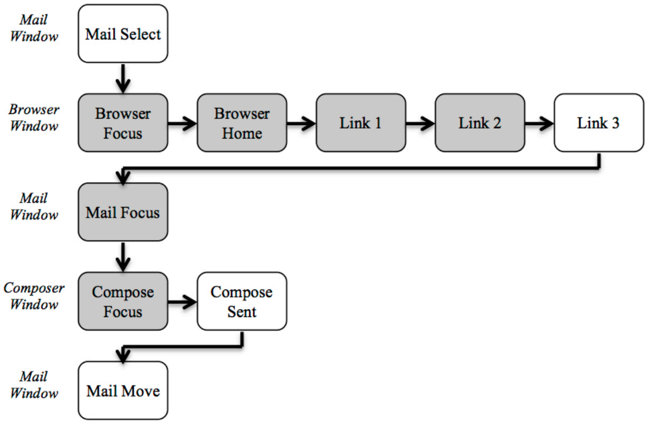


Figure 2.2. *The sequence of events of answering an email. The white blocks represent moments of low-workload and the grey ones moments of high-workload, as the participant had to remember either the model or the price of the product.*

The key characteristic of this design is that switching during the first or second link requires maintenance of the product information in working memory, while switching during the third link does not. If a chat message arrives during link 1 or 2, it is therefore better to postpone answering it until link 3 has been clicked. The delay in the email program was added to create an extra penalty for those who forgot the product information and had to return to the email and read it again. If participants switched during a delay on link 1 or link 2 (trying to take advantage of the delay time), but then forgot the product information and had to turn back to the email and read it again, they would have to wait 3 more seconds and eventually lose more time. We hoped that this would make participants more responsible when deciding to switch while they had to retain information on their working memory (high-workload moments).

Figure 2.2 shows the sequence of steps in the mail task with an indication whether or not information needed to be retained at that point. All events involve a mouse click, except “Compose Type”, where a message has to be typed in the message window, and “Link Request 1-3”, where the participant has to wait for the appearance of a webpage. There are moments of high-workload (grey boxes), where the participant’s working memory contained either the product description or the product type, and moments of low-workload (white boxes). The chat messages appearances were equally distributed in high and low workload moments.

The second manipulation we introduced concerned the difficulty of the chat messages. In the easy chat condition, participants were asked questions about movies (“Have you seen the movie…?), which required recognition and a yes/no answer. In the difficult chat condition participants were asked for a favorite book, artist, cd etc. (“What is your favorite…?”), which required recall and are open-ended questions. One third of the questions on each condition were follow-up questions related to the previous question or answer, giving the chat-task a more realistic form and making participants pay more attention to the questions. In the easy condition the follow-ups were asking “Did you like it?” if he had answered “yes” in the previous question or “Do you want to see it?” if he had answered “no”. In the difficult condition the follow-up question was “What is your least favorite?”, so the participant had to remember what the first question asked. Each participant had to complete four blocks: Delay/Difficult, Delay/Easy, No Delay/Difficult, No Delay/Easy. The goal of the difficulty manipulation was to see whether difficult questions would lead to additional disruptions in the primary task, possibly encouraging participants to avoid switching during high workload.

The blocks appeared on counter-balanced order. Each block was completed after the participant had answered 24 chat questions (8 of them were follow-up questions), so the duration of the experiment depended on the participant’s choices. However all the participants completed the four blocks in roughly one hour. Participants were instructed to give equal priority to both tasks.

RESULTS

The difficulty of the chat task variation (Difficult and Easy conditions) did not produce any significant results in either time or switching behavior of the participants, so we will collapse the data over this condition and only analyze the effect of browser delay (Delay and No Delay conditions).

We analyzed participants’ switching behavior by counting the number of switches to the chat-task after each mail-task event. We included in the analysis only the chat prompts that appeared at a high-workload point, because those were the ones that created interference to the participants. The majority of chat prompts that appeared on low-workload moments were answered immediately or in the next step (if it was also low-workload point). The proportion of switches made by all participants after each event are shown in Figure 2.3.

During the browser delay that occurs in the Delay conditions (events “Link Request 1” and “Link Request 2”) participants were tempted

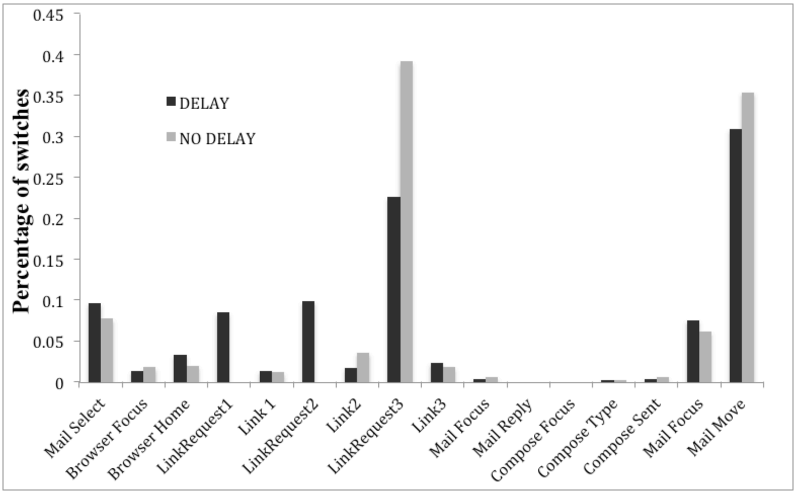


Figure 2.3. *Percentage of switches to the chat task after a mail-task event for the Delay and No Delay conditions.*

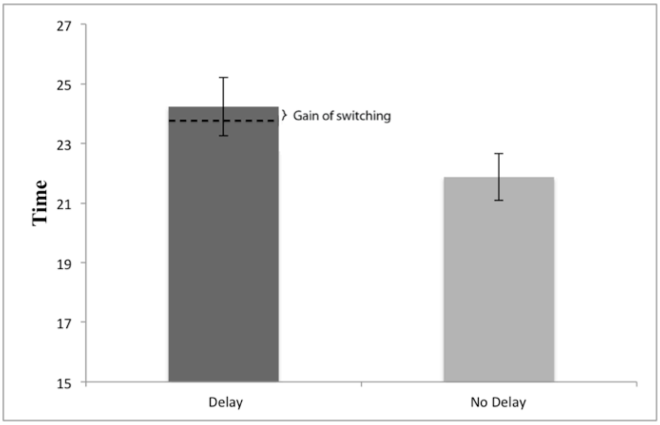


Figure 2.4. *Average time per email in the mail-task for the "Delay" and "No Delay" conditions. The line on the Delay bar shows the time gained due to switching during the browser delays.*

to switch to the chat-task, even though it is a moment of high workload. In contrast, there was only one change on those two events in the No Delay conditions. The percentage of switches on high workload points was higher in the Delay conditions (27.2%, SE 5%) than in the No Delay conditions (9.5%, SE 3.2%). A paired t-test on arcsin transformed proportions shows this difference is significant, $t(20) = 4.43$, $p < 0.001$, d=1.12.

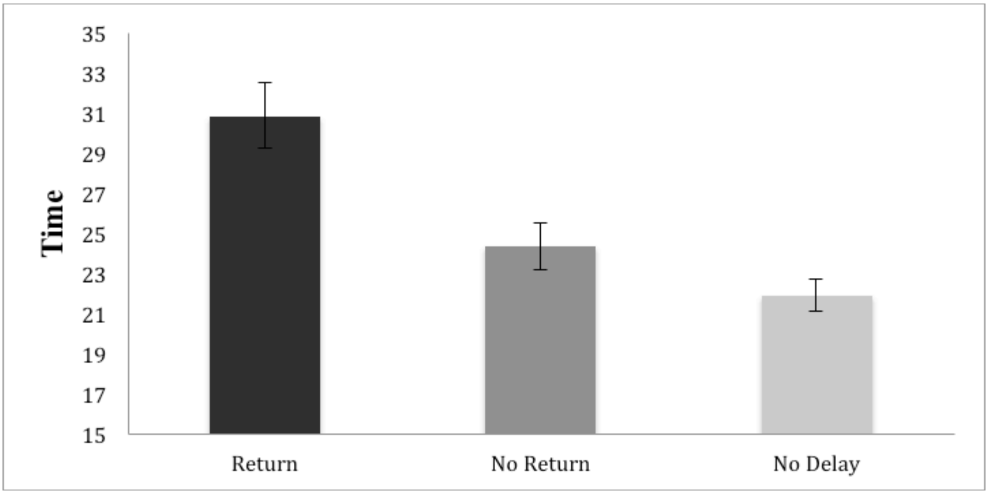


Figure 2.5. *Average time per email in the mail-task for the situations where the participants returned to the mail window to read the email information again after a bad switch ("Return"), for the situations where they didn't return after a bad switch ("No Return") and the "No Delay" situations, where there were not bad switches.*

Closer inspection of the data revealed some individual differences: all but 4 subjects made at least one switch to the chat task during a high workload pause, but some more than others. However, there was no evidence of any change in switching behavior over the course of the experiment, so no evidence for learning.

Participants switched at high-workload points in the Delay conditions, but did that affect their performance? To measure their performance we analyzed the average mail-task time, not including the browser delays in the Delay conditions and the chat-task time in both conditions. In the Delay condition participants spend on average 24.2 seconds on each mail (SE=0.87), while in the No Delay condition they needed only 21.9 seconds (SE=0.75). A paired t-test shows this difference is significant, $t(20) = 4.45$, $p < 0.001$, d=0.69. Participants answered on average 25.1 emails in the Delay conditions and 25.2 in the No Delay conditions, which is not significantly different, and they switched to the chat task approximately once during every email.

The participants' decision to switch at high-workload points during the Delay conditions made them significantly slower in the primary task (the mail-task). However, since they were going to answer the chat message at some point, choosing to do so during a forced pause in the primary task and not later could mean that they used the browser delay time productively. Since an amount of time is required to answer the chat message, why not do it during the delay and gain 3 seconds? To determine whether switching during moments of high workload had indeed a negative effect on the participants' performance, we compared the time they lost due to switching (which is the time difference between the Delay and No Delay conditions) with the gain they had from answering chat messages during the delays.

The time that participants lost because of switching on high workload moments is the mail-task time of the Delay conditions (24.2 sec) minus the mail-task time on the No Delay conditions (21.9 sec), which is 2.3 seconds. Participants made a total of 178 switches to the chat-task during high-workload moments. We included only the switches that occurred during the first high-workload moments (link request 1 and link request 2), because there was the delay variation. The time they gained is 3 seconds per switch, which is, given that a total of 1055 emails were answered in the Delay conditions, 0.51 seconds per email (which is demonstrated by the line in the Delay bar in Figure 2.4). These results show that participants lost more time (2.3 seconds) than they gained (0.51 seconds). A paired t-test shows the difference is still significant, *t*(20) = 3.49, p = 0.002, d=0.54, confirming that switching on high-workload points has a negative effect on performance.

After 97 out of the 178 switches (55%) that occurred during high-workload moments, participants had to return back to the mail window and read the product model again, whereas in the other 96 high-workload switches, they were able to recall the information and didn't have to return. The analysis done by using linear mixed-effects models showed that participants were significantly slower (*ß*=3.11, t=3.59 p=0.0002) in the mail-task for the trials in which they forgot the information and had to go back (column "Return" in Figure 2.5) than the trials in which they memorized the information (column "No Return" in Figure 2.5). The delay of the mail window was removed for the second time they returned to read the email. Still, even if they didn't forget the information, they were significantly slower from the No Delay conditions (*ß*=11.01, t=8.84, p=0.0001).

DISCUSSION

Do people switch tasks to optimize utility, or do they change tasks because they have available cognitive resources? The results of the experiment favor the latter explanation. If 3-second pauses are introduced during periods of high memory load, people tend to switch tasks even though this leads to overall decreases in performance. The extra time cost is partly incurred because participants forgot information and had to reread it. But even in cases in which participants did manage to remember the information, they were still worse off than when they delayed switching until memory load was zero.

The results are not consistent with the theory of utility maximization. A utility account would also predict that switch behavior improves as people discover utility values. However, no learning is found in the data. The results do agree with threaded cognition's "greedy" theory (Salvucci & Taatgen 2011), which states that people will switch to a task that is waiting as soon as the resources for it are available. Salvucci and Taatgen (2011), following Altmann and Trafton (2007) and Borst, Taatgen and van Rijn (2010) explain why switching during working-memory load decreases performance: when people return to the primary task after a switch, they have to restore their working memory. Therefore, when participants choose to switch tasks when their working memory contains information important for the task (as they did in the Delay condition) they need more time to restore that, compared to switching when their working memory contains nothing vital (as they did in the No Delay condition).

One aspect that the resource availability account cannot fully explain is individual differences, in particular individuals who never switch during the pauses, and individuals who only do this a few times. According to threaded cognition, when people are doing the experiment they create a goal for the mail task. But as soon as the chat-window indicates there is a message, they create a parallel chat goal. The chat goal will compete for resources with the mail goal, but as long as the mail goal occupies most resources, the chat goal will not be able to interrupt. That is why almost no switches occur during the high-load period in the control condition. However, when there is a pause in the browser, all resources except working memory are available, and therefore people switch to the chat goal. There are some possible explanations for why people do not switch. If they are actively rehearsing information during the pause their resources may be sufficiently occupied so that they do not switch. Alternatively, non-switching may be a strategic choice that people have picked up in other multitasking situations. In that case there is a tendency to switch goals, which is then overridden by an explicit strategy (like dieters consciously suppressing the tendency to eat available food).

Whichever explanation holds, it is not a strategy that develops during the experiment, because otherwise we would see effects of learning. This study by itself is not enough to support any of these explanations. However, the results are in line with observational studies in which unproductive interruptions are common, and where people try, with various levels of success, to control themselves.

PRACTICAL IMPLICATIONS

The results of our experiment suggest that a delay in the primary task is a strong trigger for switching to a secondary task. Given that the participants would switch to the secondary task at some point, doing so during a delay seems like a good decision. However, the results show that switching tasks when working memory has to sustain information is detrimental for the performance on the primary task. Participants either forget the information and have to reread it, or remember the information but still need more time to finish the primary task than when they switched at low-workload points.

This research underlines the importance of cognitive task analysis in design. A recommendation is to avoid delays during high-workload moments in task execution. If delays cannot be avoided by hardware or software means, it is better to insert delays at low-workload moments. A second recommendation is to discourage switching. That can be accomplished by making potentially distracting tasks less visible, making it less likely that the user would want to pursue them. Another option for discouraging switching is to keep the user engaged during the delay, preferably with something that helps retaining the information in working memory. Finally, given that not all participants made bad switches in the delay condition, it appears people can be trained on proper switching behavior, and it might we worthwhile to investigate this.

CHAPTER 3

Distracted by cat videos?

How the required cognitive resources affect distractibility.

ABSTRACT

Two experiments were performed, one involving problem solving and one involving visual search. In each experiment, a main task was shown together with a visual distractor (video). Participants' eye-movements were tracked to measure distraction. Results showed that participants were more distracted as the problem-solving task became harder but less distracted as the visual search task became harder. A possible explanation is that the visual system becomes less available with a harder visual task, but more available with a harder problem-solving task, where more 'thinking' and less vision is involved. This suggests that resource availability is an important factor in self-interruptions.

INTROCUCTION

In everyday life people constantly choose to stop their current task and engage in a different task. Office employees interrupt their work to check their emails, students stop studying to chat online, researchers leave papers unfinished to talk with coworkers. These interruptions are purely internally driven and are therefore called self-interruptions.

Mark, Gonzalez and Harris observed information workers and found that 52% of the interruptions were internally driven (Mark, Gonzalez, & Mark, 2005). Information workers engage in secondary tasks every 3 minutes, with 49% of these interruptions being self-interruptions (Gonzalez & Mark, 2004). Czerwinski, Horvitz and Wilhite (2004) analyzed a week-long multitasking diary of another group of information workers and found that 40% of the interruptions were internally-driven. In a study with 279 middle-school, high-school, and university students, it was shown that they disrupted their studying on average every six minutes for texting or social media (Rosen, Carrier, & Cheever, 2013). Finally, an online experiment (Gould, Cox, & Brumby, 2013) revealed that 80% of the participants switched from the routine data-entry task they had to perform for one hour at least once, despite the warning that switching would result in reward reduction.

Although there are many experimental studies of interruptions, most of them focus on external interruptions (e.g. Brumby, Cox, Back, & Gould, 2013; Hodgetts & Jones, 2003; 2006; Monk, Boehm-Davis, & Trafton, 2004; Monk, Trafton, & Boehm-Davis, 2008) and their negative effects, such as time costs (e.g. Monk et al. 2008) or higher error rate (e.g. Brumby et al., 2013). However, according to observational studies, self-interruptions are just as common as external interruptions (Czerwinskt, Horvitz, & Wilhite, 2004; Gonzalez & Mark, 2004; Gould et al., 2013; Mark et al., 2005; Rosen et al. 2013). Although it is hard to create an experimental setup in which people are completely free to self-interrupt, the few studies that use such a setup show the negative effects of self-interruptions (e.g., Katidioti & Taatgen, 2014; Katidioti, Borst, & Taatgen, 2014). To help users manage those interruptions and achieve optimal performance in everyday tasks, it is crucial to acquire more knowledge on self-interruptions. Specifically, to manage the impact of self-interruptions, we need better insight in the internal causes behind self-interruptions.

In this study we will focus on one type of self-interruption: distractions (i.e when attention is drawn from the current task one is performing). Getting distracted can be the first step to making a self-interruption. Distractibility can be affected by many factors, such as the type of task one is currently performing and the type of distractor. In the current study we investigated the effects of type-of-task (defined as the cognitive resources required), in combination with task difficulty. The type-of-task conditions we used were visual (visual search task) and working memory (problem-solving task).

Cognitive resource availability

Wickens' (2002) multiple resources theory explains that there is more interference when combining two tasks that share the same cognitive resource (e.g. two visual tasks) compared to two tasks using different cognitive resources (e.g., one visual and one auditory task). Salvucci and Taatgen's (2011) threaded cognition theory also supports the idea that tasks using the same cognitive resources are harder to combine than tasks using different cognitive resources. There are a number of studies that confirm multiple resources theory and threaded cognition by showing that participants show worse performance when combining two tasks using the same cognitive resource (e.g., Borst, Taatgen, & van Rijn, 2010; see Salvucci & Taatgen, 2011; Wickens, 2002 for reviews). Furthermore, when allowed to choose, participants prefer to combine tasks that do not share any cognitive resources over combining tasks that share a cognitive resource (Nijboer, Taatgen, Brands, Borst, & van Rijn, 2013).

In addition, availability of cognitive resources seems to be an important reason for someone to self-interrupt. While playing a memory game, participants chose to self-interrupt mostly after they had found a match between two cards, and the number of items in their working memory therefore decreased (Katidioti et al., 2014. In another study when participants were given the freedom to choose when to self-interrupt, they demonstrated rational behavior and chose to switch tasks on low-workload moments (i.e., when their working memory was not occupied) (Salvucci & Bogunovich, 2010). However, this rationality disappeared when they were forced to wait 3 seconds for an Internet browser to load (Katidioti & Taatgen, 2014). As soon as the idle loading time made cognitive resources for the secondary task available, participants chose to switch to the secondary task, even though that made their performance worse. These studies seem to suggest that people switch when cognitive resources for the new task (e.g., vision, working memory) become available.

Current study

Although previous studies have shown that resource use determines how well multiple tasks can be performed together (Salvucci & Taatgen, 2011; Wickens, 2002), it is unclear if resource availability also plays a role in getting distracted by an irrelevant task. With the current study we try to expand the theories of cognitive resource availability into the field of distractibility and consequently self-interruption. Our hypothesis is that distractibility depends on the resources the task shares – or does not share – with the distractor. A distraction that is related to a certain resource (e.g., a video, using visual resources), will make people more distracted in tasks in which this resource is available.

In order to test this hypothesis, we used a visual and a problem-solving task as the main tasks of two experiments, while the distractor was purely visual (a silent video). A complicating factor is that cognitive resource availability often depends on task difficulty. On the one hand, if the main task shares a resource with the distracting task (e.g., a visual-search task and a visual distractor), the harder the main task gets – and thus uses the visual resource more – the less easy it will be for the distractor to intrude. On the other hand, if the main task mainly uses a different resource (e.g., a problem-solving task and a visual distractor), it will use the visual resource less – because more time is spent ‘thinking’ – when it gets harder, giving the distractor a greater opportunity for intruding. Therefore, the two tasks had three difficulty levels each. Our prediction was that people would be less distracted by the video as the visual task gets harder and visual resources are more engaged. On the other hand, as problem solving gets harder the visual aspects of the task will change slower, making visual resources more available and leading participants to be more distracted by the video.

These predictions are in line with results from visual perception studies. According to Lavie’s (1995) load theory, people are less likely to reject distractors as task difficulty increases in visual tasks. Later experiments confirmed that effect and also found the opposite effect for tasks using working memory (see Lavie, 2010 for a review). While Lavie’s studies investigate distractors that form part of the main task display, the study presented here uses distractors that are incidental to the main task. Both our tasks and our distractors resemble everyday life situations and for that reason the results can be generalized to working environments. These allow us to make suggestions about optimizing software and design of workplaces.

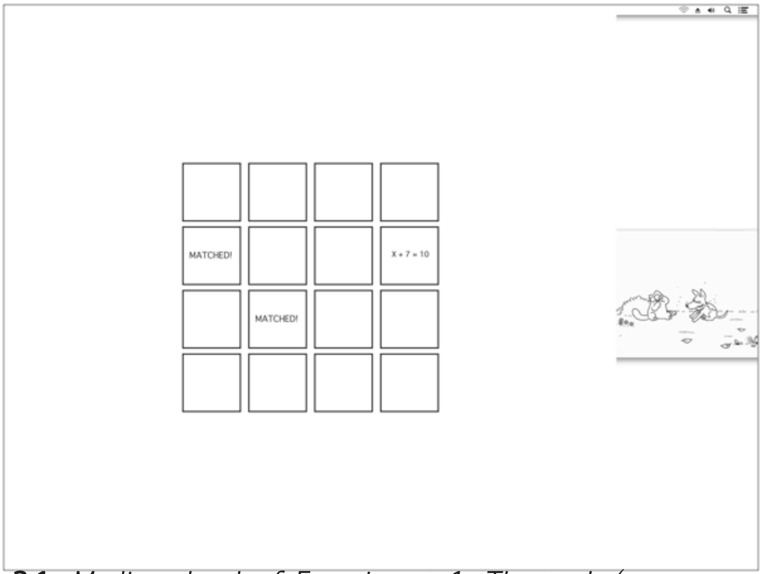


Figure 3.1. *Medium level of Experiment 1. The task (memory game) is on the left and the distractor (video) on the right*

EXPERIMENT 1

The main task of this experiment (also used in Katidioti et al., 2014) was a variation of the game known as Concentration or Memory, replacing images with equations (Figure 3.1). Participants had to click on a card with the left mouse button, mentally solve the equation that appeared on the card, remember the value of *X* and continue by clicking on another card. In this version of the game, cards are opened one at a time instead of in pairs and a match is made when the opened card matches the previously opened card (the word "MATCHED" appeared on the back of these cards and they could not be clicked again). This task was chosen because it requires problem-solving and working memory, two aspects used by many everyday tasks, such as homework, data entry, programming etc.

In our instantiation (similar to Anderson, Fincham, Schneider & Yang, 2012) there are 16 cards (8 pairs) with equations on them, arranged in a 4 by 4 matrix (Figure 3.1). In the equations, X was the unknown value and was always an integer from 2 to 9. There were three difficulty levels:

- **Easy**: $a + b = X$, where a and b were integers from 0 to 9
- **Medium**: $X + a = b$, where a was an integer from 1 to 9.
- **Hard**: $a*X + b = c$, where a and b were integers from 1 to 9.

Two cards are said to match when they have the same value for

X. For example the cards “$2 * X + 2 = 12$” and “$3 * X + 4 = 19$” are a match, since X=5 in both of them.

In addition to the experimental window with the memory game, there was also a video playing constantly on the screen (Figure 3.1). This mute video was compiled of 37 clips of “Simon’s Cat”[1], with a total duration of 50 minutes. The video was a black and white cartoon with simple line drawings and simple stories that are easy to follow at any moment. This distractor was chosen because it closely resembles everyday distractors, such as televisions in the working area, street advertisements or a car DVD player while driving, even a window or an open office door.

Apparatus and setup

Participants were tested individually in a windowless room, on a LCD monitor (resolution: 1600x1200 pixels, density: 64 pixels/inch), using a chin-rest. The eyetracker was an Eyelink 1000 from SR Research. Eye fixations were measured with a sample rate of 250 Hz. The dimensions of the experimental window were 1200x1300 pixels, while the video was 300 pixels wide. Half of the participants completed the experiment with the experimental window on the left part of the screen (pixels 0 to 1300) (as in Figure 3.1) and the other half with the experimental window on the right part of the screen (pixels 300 to 1600).

Procedure

Participants first practiced one simple memory game (match the numbers) and one Easy level memory game. The real experiment lasted 45 minutes, one 15 minute-block for each difficulty level, with as many memory games as they could fit. When the 15 minutes were over, participants finished the memory game they were playing at the moment and could take a break. The order of appearance of each difficulty level was counterbalanced, so that all possible orders would appear equally often. Participants were informed beforehand that they would complete three blocks of 15 minutes and were instructed to maximize accuracy. There were no instructions with respect to the video. No explicit feedback was given in this task.

Participants

25 participants (14 female) participated in the experiment. One

---

[1] http://www.simonscat.com/
Simon’s Cat©

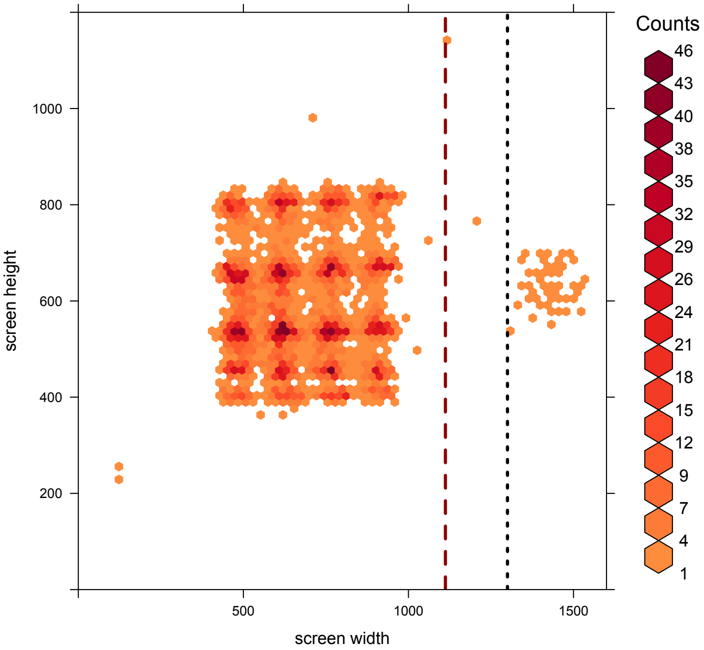


Figure 3.2. *All fixations during the whole duration of the experiment for one participant. The dotted black line shows the end of the experimental window (and the beginning of the video window) and the red dashed line represents the halfway point between the video and the edge of the memory game that was used as the limit to distinguish fixations to the video from fixations to the memory game*

male participant was removed because of malfunction of the eyetracker. The remaining 24 participants had a mean age of 23.76. All participants had normal or corrected-to-normal vision, gave informed consent for their participation and received monetary compensation of 10 euros.

Results

Two participants were not distracted at all (they had zero fixations on the video) and were removed from further analysis, since their data would not provide insight into task distraction. For each memory game, the time between the last card match in the game and the beginning of the next memory game was removed from the analysis, as that time could be used to look at the video without it constituting distraction.

Table 3.1 shows that our difficulty manipulation worked as expected. Participants were the fastest to complete a memory game in the Easy

	Easy	Medium	Hard
Average time per memory game (sec)	48.02 (2.46)	87.15 (6.32)	240.00 (19.11)
# of memory games completed in 15 minutes	19.09 (0.86)	11.23 (0.63)	4.68 (0.37)
Average # of revisits	8.51 (1.12)	12.37 (1.4)	23.64 (3.14)

Table 3.1. *Task performance as a function of difficulty in Experiment 1 (parentheses show standard errors)*

level and the slowest in the Hard level. All levels were significantly different from one another according to a repeated-measures ANOVA ($F(2,42)=101.3$, $p<.001$, $\eta_p^2= 0.83$) and Follow-up t-tests confirmed that (all $ps<.001$).

Fixations

We plotted the fixations of each participant to ensure that our measurements were correct. Figure 3.2 shows the fixations of a typical participant. Fixations to the right of the red line were considered distractions.

A second measure of performance is the average number of times participants reopened a card they had already seen (number of revisits). As Table 3.1 shows, the number of revisits was also affected by the difficulty level, with participants having on average the fewest revisits in the Easy level and the most in the Hard level. The levels were significantly different from one another according to a repeated-measures ANOVA ($F(2,42)=18.34$, $p<.001$, $\eta_p^2= 0.47$) and follow-up t-tests (all $ps<.005$).

We defined "initial fixations" as the number of times participants moved their eyes from the memory game to the video, thus the number of times they got distracted. We defined "time spent watching the video" as the whole duration they fixated on the video area. There were large individual differences in the number of initial fixations (getting distracted from 3 times to 122 times during the 45-minute experiment), and the time spent watching the video (from less than 2 seconds to more than 30

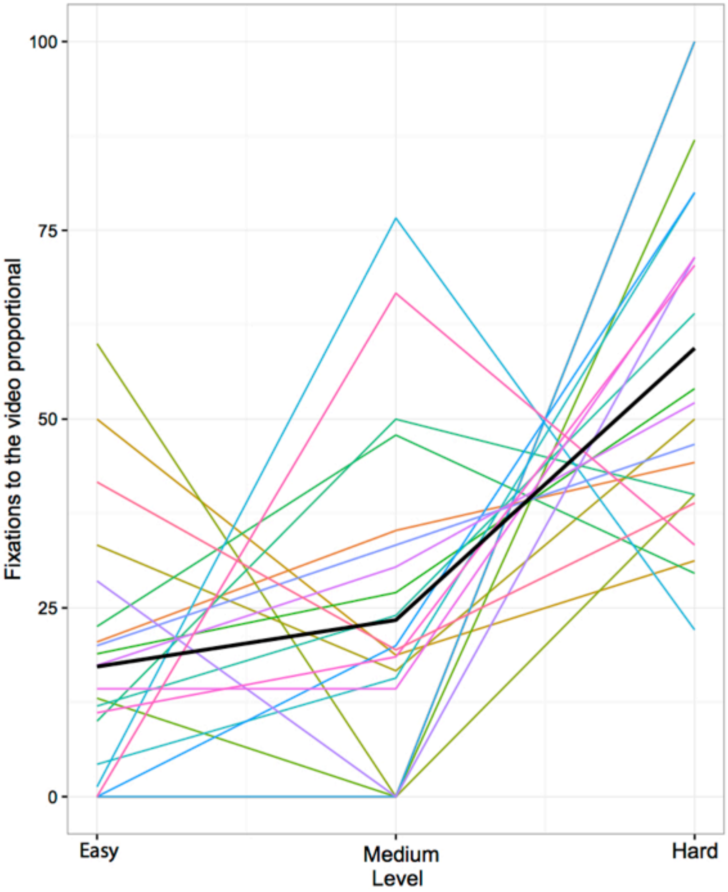


Figure 3.3. I*nitial fixations to the video for each participant in Experiment 1 (proportionally for each level). The average is represented by the black line.*

seconds during the 45 minutes). Consequently, an analysis on the average number of initial fixations or the average time spent watching the video was not possible. We used a linear mixed effects model with a Poisson distribution to compare the number of fixations to the video for each level. Results showed that the number of fixations differed significantly between each pair of conditions: Easy-Medium: Easy-Medium (*ß*=0.68, z=5.69, p<.001), Easy-Hard (*ß*=1.11, z=9.85, p<.001) and Medium-Hard (*ß*=0.43, z=0.09, p<.001).

We converted the data into proportional data per level for each participant, meaning that out of all the initial fixations the participant made to the video, we computed the percentage that occurred in the

Easy, Medium and Hard levels. The proportion of fixations to the video was lowest in the Easy level (17.23%), higher in the Medium level (23.39%) and the highest in the Hard level (59.39%). Figure 3.3 shows the data for each participant (and the average) for the initial fixations to the video, proportional for each level. As the average (black line) demonstrates, as the task got harder, there were more fixations to the video. Furthermore, we calculated the regression slopes for the initial fixations to the video (proportional per level), of which the vast majority were positive (average slope 21.08, which is significantly larger than zero $t(21)=5.51$, $p<.001$ and $d=1.18$).

Likewise, participants watched the video for proportionally less time in the Easy level (13.29%), more in the Medium level (20.65%) and even more in the Hard level (64.18%). The data had a similar pattern to Figure 3.3 and the regression slopes for time spent watching the video were also mostly positive (average slope 25.45, $t(21)=6.74$, $p<.001$ and $d=1.44$). There was a high positive correlation between the number of fixations to the video and the time spent watching the video with $r=0.91$ and $p<.001$. The linear mixed effects model revealed significant difference between all levels in the time spent watching the video: Easy-Medium (*ß*=0.66, $z=119.73$, $p<.001$), Easy-Hard (*ß*=0.83, $z=214.12$, $p<.001$) and Medium-Hard (*ß*=1.03, $z=4.04$, $p<.001$).

Performance

Task difficulty clearly affected distractibility, but that was not reflected in task performance. There was no correlation between distractibility measures (initial fixations to the video and time spent watching the video) and performance measures (number of memory games completed per block, time needed to complete a memory game and number of revisits). Therefore, participants did not become slower or forget the cards more easily when they were more distracted.

Discussion of Experiment 1

In this experiment we examined how much people were distracted by a visual distractor during a problem-solving task. As the task got harder, participants became slower (Table 3.1). As the participants – and therefore the task – became slower, the visual aspects of the task also changed more slowly (i.e., when the task is more difficult, it takes longer to solve an equation and move to a new card), hence making visual resources more available. Results showed that participants were

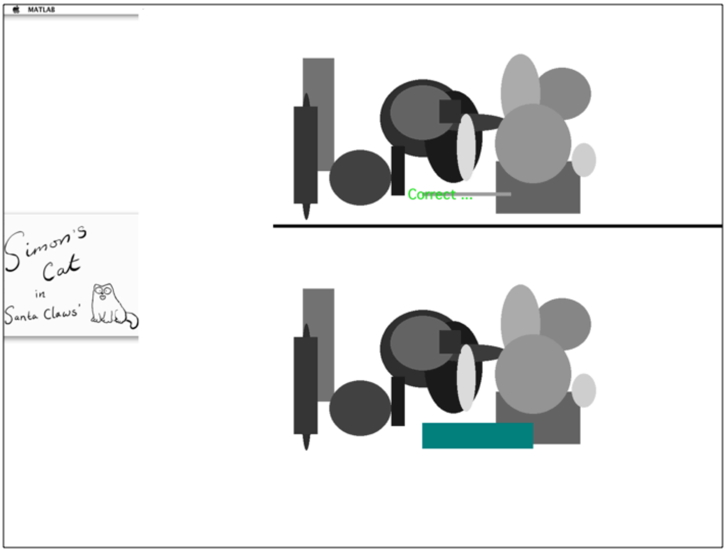


Figure 3.4. *Medium level of the main task of Experiment 2, during the feedback moment.*

more distracted by the video as the task became harder and the visual resources more available. These data support our "available resources" hypothesis of distraction.

EXPERIMENT 2

The goal of this task was to find the difference between two almost identical images. The images were compiled of randomly drawn shapes (circles, ovals, squares and rectangles), colored with randomly chosen grey and black color shades. The experimental window was divided into a top part and a bottom part, with a black line separating them (Figure 3.4). Both parts were identical, except that one of the shapes in the bottom part was bigger than its corresponding shape in the top part. Participants had to locate that shape and click on it with the left mouse button. After the participant had made their choice, feedback was given by highlighting the correct shape with a blue color, accompanied by the word "Correct" or "Wrong" shown in the top part of the screen (as seen in Figure 3.4). After this feedback, the new trial began. As in Experiment 1, there were three levels of difficulty, which were determined by the size and number of the shapes. We chose this task because it uses visual search, which is used in many everyday tasks, such as proof-reading a document, graphic or architectural design, etc. The video distractor was used in the same

	Easy	Medium	Hard
Average responce time (sec)	3.26 (0.24)	9.25 (0.74)	18.4 (1.46)
# of trials completed in 15 minutes	159.92 (5.09)	82.32 (3.9)	49 (3.21)
% of wrong answers	1.5% (0.27%)	1.82% (0.45%)	3.34% (0.58%)

Table 3.2. *Task performance as a function of difficulty in Experiment 2 (parentheses show standard errors)*

way as in Experiment 1.

Procedure

Before the main task, participants practiced the task for 4.5 minutes (1.5 minutes for each level). The experiment lasted 45 minutes, one 15 minute-block for each level, with as many trials as participants could fit in that time. When the 15 minutes had passed, participants finished the trial they were playing at the moment, after which they could take a break. The order of appearance of each difficulty level was counterbalanced, so that all possible orders would appear equally often. Participants were informed beforehand that they would complete three blocks of 15 minutes and were instructed to maximize accuracy. There were no instructions with respect to the video. After each trial a “correct” or “wrong” feedback was provided.

Apparatus and setup

Identical to Experiment 1.

Participants

25 participants (15 female), with a mean age of 24.76, participated in the experiment. All participants had normal or corrected-to-normal vision, gave informed consent for their participation and received monetary compensation of 10 euros.

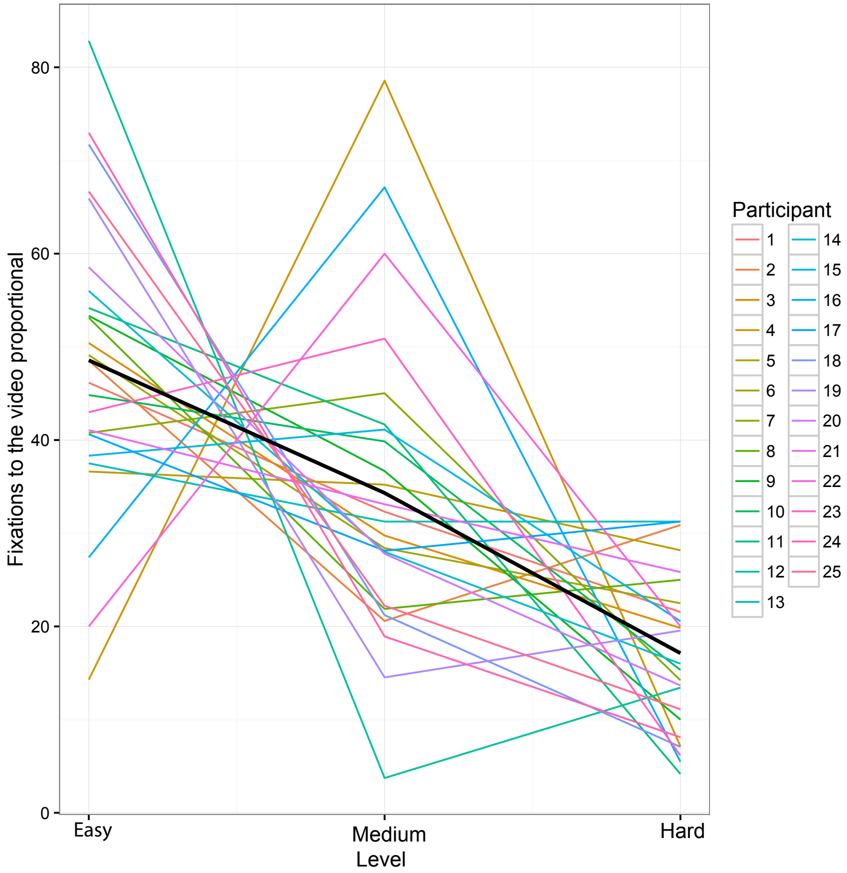


Figure 3.5. *Initial fixations to the video for each participant in Experiment 2 (proportionally for each level). The average is represented with a black line.*

Results

Table 3.2 confirms that our difficulty manipulation worked. Participants found the difference the fastest in the Easy level and the slowest in the Hard level, with all levels being significantly different from one another according to a repeated-measures ANOVA ($F(2,48)=115.3$, $p<.001$, $\eta_p^2=0.83$). There was also a difference between difficulty levels in the percentage of wrong answers, with participants making the most mistakes in the Hard level and the fewest in the Easy level.

We analyzed the fixations of each participant in the same way as in Experiment 1 (see Figure 3.2) and used the same method to separate the fixations to the video from the fixations to the task. We included the feedback time (1 second) in this analysis, since it is an important part of the game during which looking at the video can be classified as a distraction[2]. However this low-workload moment probably provided participants with a larger opportunity to get distracted than in Experiment 1. There were again large individual differences in the initial fixations to the video and the time spent watching the video. Once more, we used a linear mixed effects model with Poisson distribution to compare the number of fixations to the video for each level. Results showed that all levels had a significantly different number of fixations: Easy-Medium (ß=-0.47, z=-10.5, p<.001), Easy-Hard (ß=-1.08, z=-19.65, p<.001) and Medium-Hard (ß=-0.61, z=-10.37, p<.001).

The proportion of initial fixations to the video (i.e. how many times participants looked at the video) followed the opposite pattern from Experiment 1: it was the highest in the Easy level (48.55%), lower in the Medium level (34.32%) and the lowest in the Hard level (17.13%). Figure 3.5 shows how many initial fixations to the video each participant made, proportional for each level. As the average (black line) demonstrates, as the task became more difficult, participants made fewer fixations to the video. We also calculated the regression slopes for the initial fixations to the video (proportional per level), which were all negative (average slope -15.71, t(24)=-7.98, p<.001, d=1.6).

Participants watched the video for proportionally more time in the Easy level (48.29%), less in the Medium level (34.26%) and the least in the Hard level (17.45%). The data exhibited a pattern similar to Figure 3.5 and the regression slopes for time spent watching the video were mostly negative (average slope -15.41, t(24)=-6.41, p<.001, d=1.28). There was a positive correlation between the number of fixations to the video and the time spent watching the video (per level), r=0.9, p<.001. In addition, the linear mixed effects model revealed a significant difference between all levels for the time spent watching the video: Easy-Medium (ß =-0.38, z=-285.4, p<.001), Easy-Hard (ß =-1.21, z=-682.1, p<.001) and Medium-Hard (ß=-0.83, z=-444.2, p<.001).

Discussion of Experiment 2

In Experiment 2 as the task became harder, the visual resources were more engaged. Results showed that participants were less distracted by the video as the task became harder and the visual resources became less available. In other words, these data again support our available resources hypothesis.

---

[2] However, removing the feedback time did not alter the results

GENERAL DISCUSSION

With the current study we extended the ideas of multiple resources theory (Wickens, 2002) and threaded cognition (Salvucci & Taatgen, 2011) to distractibility and hence self-interruption. We conducted two experiments, in which participants had to work on a task while a video was playing next to the task, causing a possible visual distraction. The two experiments required different cognitive resources: the task of Experiment 1 required problem solving and working memory, whereas the task of Experiment 2 required visual resources.

Generally participants were not distracted by the video for a long time, but even this minimal distractibility was enough to show an effect of cognitive resource availability on distractibility, which are in line with multiple resources theory (Wickens, 2002) and threaded cognition (Salvucci & Taatgen, 2011). Fixation analysis showed that in the problem-solving/working memory task participants were more distracted as the task became harder and therefore required fewer visual resources (the visual aspects of the task changed slower). In contrast, in the visual search task participants were less distracted as the task became harder and required more visual resources. Based on these results, we recommend avoiding having moving visual distractors around (e.g., TV in the working/studying environment, open office doors etc.) when working on a task that does not occupy all visual resources. More generally, it would be prudent to avoid all distractions that use resources not used by the task one is currently performing, such as visual distractors when solving mathematical problems, auditory distractors when reading a paper etc. For example, it might be useful to block out certain auditory distractors (colleagues talking in the corridor) by occupying the auditory resources with well-known music. Open office spaces should be avoided if the job requires minimal visual resources (e.g. brainstorming, telesales etc.). Designers can use these results in order to create or improve software that tracks the user's behavior and customizes the distractions accordingly. There are a number of jobs that require the use of multiple monitors or devices and software that takes into account cognitive resources availability could minimize self-interruptions.

There are some situations in everyday life in which it is desirable to be distracted. For instance an urgent notification should be auditory if a person is working on a visual task. A safety alarm system in a factory might be visual distractor (e.g., a red light), but, if workers are occupied with high difficulty visual tasks (e.g. finding an error in a programming code), the visual alarm might not be perceived. A change in a stock price should be announced a sound if the broker is writing a document. These suggestions can be expanded to other situations that one should get immediately distracted if there is a reason, such us flying a plane, working in a control tower or a factory with many machines etc.

The results of this study provide us with insight into the mechanisms behind self-interruptions in general. Having the resources for an interruption free can lead someone to self-interrupt. Resource availability often depends on task difficulty, since an easy task uses fewer resources in general than a hard task. According to our results, a task can make people less distracted as it gets harder if it shares a resource with the distractor, whereas another task can make people more distracted as it gets harder if that makes the resources used by the distractor more available. These results can have many practical implications for improving working environments.

Limitations of this study

Firstly, our study only tests the available resources hypothesis for distractors using the visual resource. In order to generalize into the relationship between the availability of other cognitive resources and distractibility, a follow-up experiment is needed, where a different type of distractor will replace the visual distractor. That distractor should share more resources with a problem-solving task as the task gets harder and less with a visual task as the task gets harder.

Secondly, participants were distracted only for a very brief period, which is probably quite different from real-life distractions by e.g., Facebook, in which case the participant is distracted for a period of seconds to minutes.

PART 2

External vs. self-interruptions

CHAPTER 4

What happens when we switch tasks: *pupil dilation in multitasking.*

This chapter was previously published as:
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

ABSTRACT

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.

INTRODUCTION

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 minutes, while Chisholm and colleagues reported that physicians in an emergency department were interrupted on average about 50 times in 180 minutes (Chisholm, Collison, Nelson, & Corcell, 2000). Observing students in their home environments, Rosen, Carrier and Cheever's (2013) found that they studied on average 6 minutes 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, while 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 cellphone, disable email 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 to external interruptions and investigating the effects of a lag between the start of the interruption and the start of the secondary

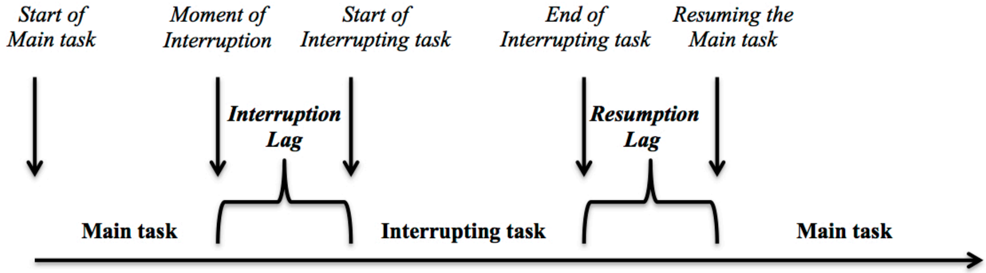


Figure 4.1. *Time course of an interruption (based on Trafton et al., 2003)*

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 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 4.1 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; Roger & Monsell, 1995; Trafton et al., 2003).

One of the major cognitive theories on interruptions is Altmann & Trafton's memory for goals theory (2002). 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.

While 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 mid-subtask and after subtasks and found that an interruption mid-subtask 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 et al. (2003) performed an experiment where participants either had an alert before the secondary task began, followed by an 8-second 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 & 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 second delay.

Monk et al. (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, since 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 thirteen 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, since 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 unrelated task when a sub-task 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 mid-subtask. 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 to 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 6 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 second 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 3 to 7 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 7th or 8th 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, since 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 1250 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, i.e. inhibit a response. Satterthwaite, Green, Myerson, Parker, Ramaratnam & 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

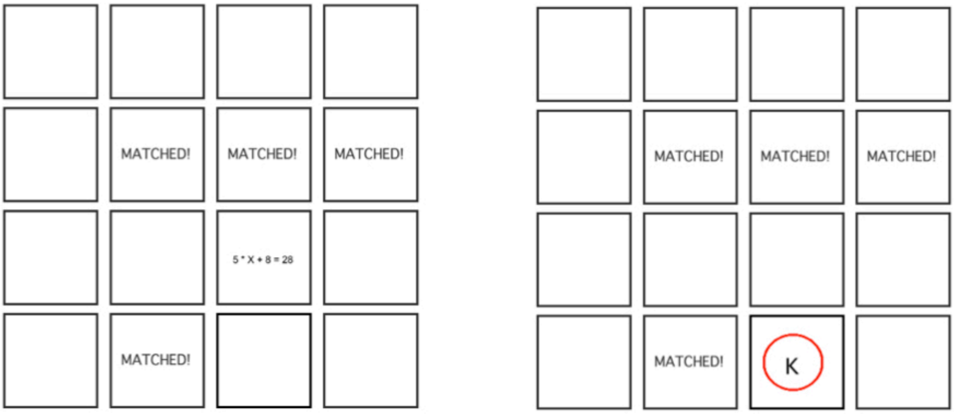
Method

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 (8 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 by 4 matrix (Figure 4.2). 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, since $X=5$ 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



4.2a **4.2b**

Figure 4.2. *The interface of the primary task (memory game) (4.2a) and the secondary task (n-back) (4.2b)*

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, 1985). 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, 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, e.g. 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 4.2a). 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.

In order 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 seconds after which the n-back task started. The n-back task lasted 15 seconds and the probability of a letter being the same as two letters before was 50%. Each letter stayed on the screen for 1.5 seconds, 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 4.2b). After the feedback, there was 0.5-second 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 etc.

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 4.2b). 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 to self-interruptions, while 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 seconds. In the Delay condition there was a 3 second 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 (3 blocks Delay/Forced, 3 blocks Delay/Voluntary, 3 blocks No Delay/Forced and 3 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 3 times within a block (in the Voluntary condition), otherwise they would get a penalty of a 30 second 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 hour). 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 minutes.

Participants

26 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 inch LCD monitor with screen resolution of 1,024 x 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 45cm 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 seconds at a fixation cross.

Measurement and preprocessing of pupillary data

Eye blinks 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 seconds before and 10 seconds after every switch and every click. Switches that had more than 20% of their data (20% of the 15 seconds 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, 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).

Results of Experiment 1

Behavioral results

To remove participants that did not actually perform the task, we set a threshold of an average time of 3 seconds per card per block. We assumed people could not solve an equation on average in less than 3 seconds per card per block. If a block had an average time of less than 3 seconds, 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

		Condition			
	Average	**Forced**	**Voluntary**	**Delay**	**No Delay**
Time per block (sec)	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)(sec)	191.11 (8.57)	187.16 (9.9)	195.28 (8.73)	192.65 (8.98)	190.6 (9.45)
Nr of revisits per block	12.61 (1.74)	13.42 (2.16)	11.78 (1.52)	12.91 (1.75)	12.36 (1.9)
Nr 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)
Nr 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)
Nr 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)

Table 4.1. *Behavioral data of Experiment 1 (mean (SE))*

was removed. Three participants and 6 blocks from 4 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 4.1 show that the differences between conditions were minor. There was no statistically significant difference amongst 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 2nd and 4th row of Table 4.1). 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=0.24 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 seconds to prepare for the secondary task. However, having a 3 second 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 seconds and in the Forced/No Delay condition 181.78 seconds, the difference of 8.67 seconds was not significant ($t(21)=0.77$, $p=0.45$, $d=0.17$).

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. Since 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 seconds and 2.98 seconds for the Voluntary/No Delay condition, with that difference not being significant ($t(21)=0.48$, $p=0.63$, $d=0.08$), 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 an FDR correction over all samples

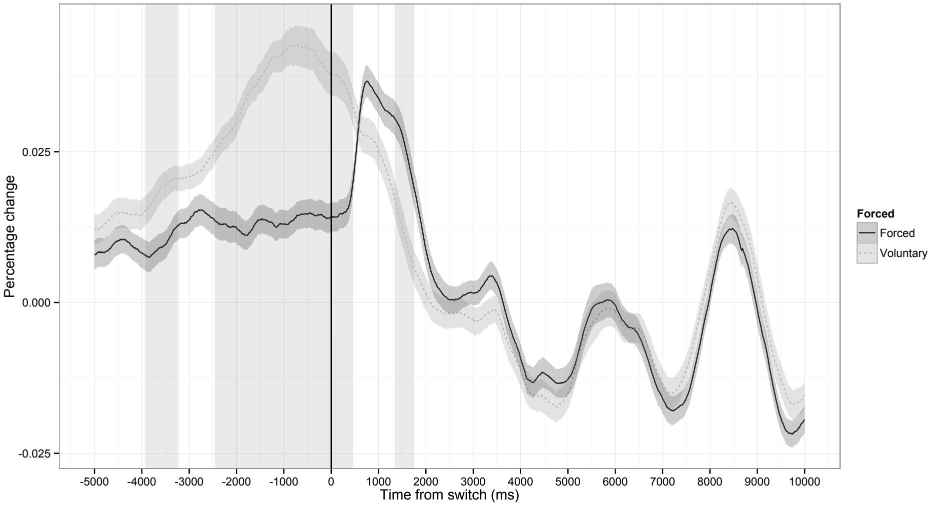


Figure 4.3. *Average pupil dilation of all participants for the Forced and Voluntary conditions of Experiment 1 around the switch (0 is the switching point from memory game to the n-back task). Marked with dark grey color are the moments where there was a statistically significant difference (p<.05, FDR-corrected) between the Forced and the Voluntary conditions. Shading indicates a standard error.*

to correct for multiple comparisons (Benjamini & Hochberg, 1995). There was no interaction between conditions in increase of pupil dilation. Figure 4.3 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 dilation for both conditions, indicating the participants' responses in the n-back task.

In Figure 4.4 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

---

[1] We did not plot the pupil dilation increase for the Forced/Delay and Forced/No Delay conditions, since whatever processes may happen there (rehearsal, preparation) will be even stronger when participants choose themselves to switch tasks, i.e. the Voluntary condition.

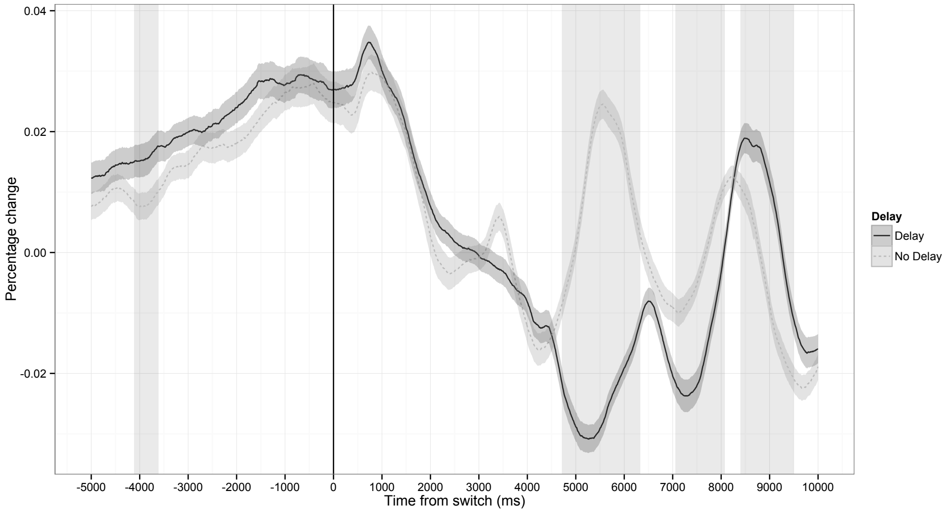


Figure 4.4. *Average pupil dilation for all participants in the Voluntary condition for the Delay and the No Delay conditions of Experiment 1 around the switch (0 is the switching point from memory game to the n-back task). Marked with dark grey color are the moments where there was a statistically significant difference (p<.05) between the Delay and the No Delay conditions. Standard error is shown with a lighter color around the line.*

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 (Figure 4.5). 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 of Experiment 1

Switching tasks created an increase in pupil dilation (Figure 4.3 and Figure 4.4). But what is the cause of that increase? As we indicated in the introduction, 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

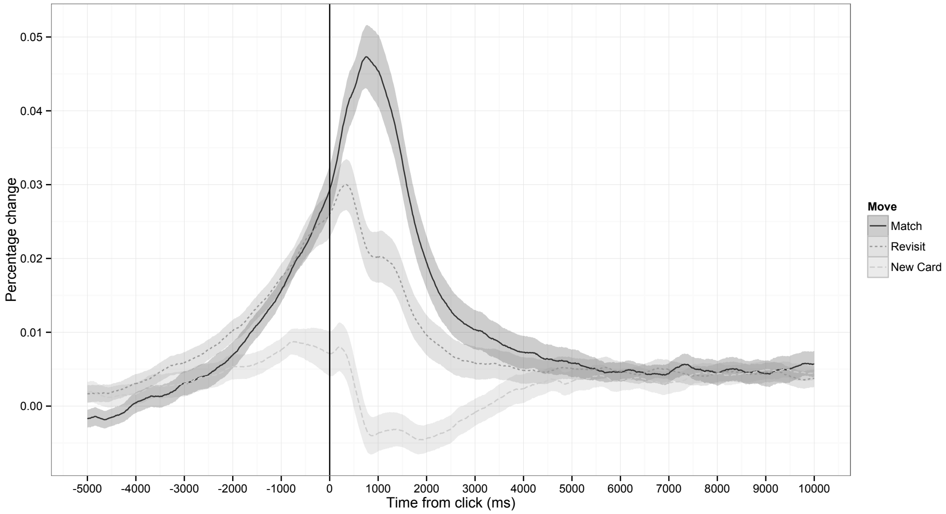


Figure 4.5. *Average pupil dilation for all participants of Experiment 1 in all conditions graphed according to what kind of move they make after 0 (0 is the time of clicking on a card). Standard error is shown with a lighter color around the line.*

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 4.4 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 (Table 4.1). 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 4.5, 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 4.3 and Figure 4.4). 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 (Figure 4.3). Making a self-interruption (Voluntary condition) created a reaction in the pupil some seconds before the switch, while being externally interrupted (Forced interruption) created a sudden peak in pupil dilation at the moment the interruption was made (assuming a one-second 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, since 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, since the Forced condition was an abrupt external interruption that happened mid-task, 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 seconds on average) to complete the primary task in the Forced condition than in the Voluntary condition (Table 4.1). Why isn't a mid-task interruption more disruptive than a post-task 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 mid-task 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 (since 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 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 7 matches in a block that could be a switching point (the 8th 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, i.e. 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 (7th) match: if there had not been 3 switches by that point, a switch occurred after the 7th match. That safe point ensured that there were going to be 3 switches per block. Since usually the switches happened after a match, we picked the 7th match (last match after which a switch could occur) as a safe point. A similar safe point was placed after the second-to-last (6th) 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

25 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 of Experiment 2

Behavioral results

Three participants and 14 blocks from 4 other participants were

		Condition			
	Average	**Forced**	**Voluntary**	**Delay**	**No Delay**
Time per block (sec)	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)(sec)	200.93 (12.96)	191.03 (11.66)	210.54 (14.56)	199.92 (12.62)	201.57 (14.13)
Nr of revisits per block	14.37 (1.99)	13.74 (1.98)	14.88 (2.21)	14.64 (2.17)	14.09 (2.07)
Nr 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)
Nr 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)
Nr 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 (sec)	2.46 (0.12)	2.34 (0.12)	2.59 (0.15)	2.41 (0.14)	2.48 (0.12)

Table 4.2. *Behavioral data of Experiment 2 (standard errors in the parenthesis)*

removed because they did not pass the threshold of 3 seconds 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 7th match) was used only in 2 blocks across all participants. The second safe point (after the 6th 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 seconds per memory game (time per block minus the time spent on the secondary task) in the Forced condition and 210.55 seconds in the Voluntary condition, which was a significant difference of 19.52 s.; t(21)= -3.53, p=0.002, 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, η_p^2=0.27). This time the Forced condition produced switches mirroring the

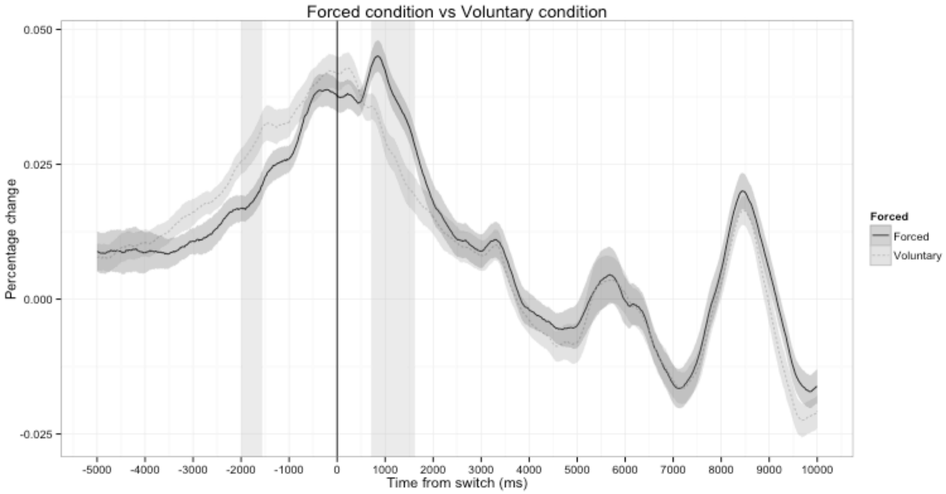


Figure 4.6. *Average pupil dilation for all participants of Experiment 2 for the Forced and Voluntary conditions around the switch (0 is the switching point from memory game to the n-back task). Marked with dark grey color are the moments where there was a statistically significant difference ($p<.05$) between the Forced and the Voluntary conditions*

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 second 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 seconds and in the Forced/No Delay condition was 182.21 seconds, with that difference not being significant ($t(21)=1.86$, $p=0.08$, $d=0.28$).

Analyzing the resumption lag was possible for all conditions in this experiment (last line of Table 4.2). 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 ($F(1,20)=5.4$, $p=0.03$, $\eta_p^2=0.21$). There was no significant difference between the Delay and the No Delay conditions and no interaction[2].

---

[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.

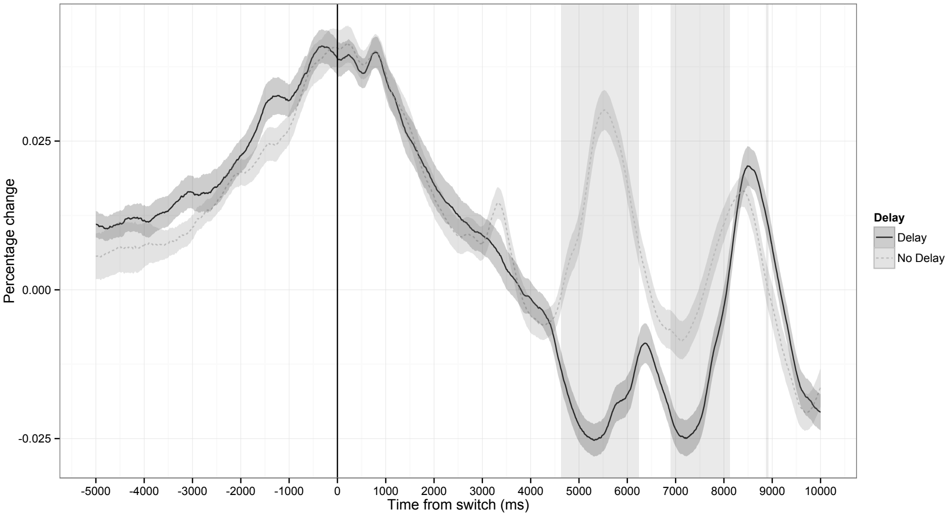


Figure 4.7. *Average pupil dilation for all participants in the Voluntary condition of the 2nd experiment the Delay and the No Delay conditions around the switch (0 is the switching point from memory game to the n-back task). Marked with dark grey color are the moments where there was a statistically significant difference (p<.05) between the Delay and the No Delay conditions*

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 4.6 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 4.7 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 seconds before the switch, but as we will explain later, we believe this is coincidental.

Discussion of Experiment 2

There was a difference in the increase of pupil dilation between the Forced and the Voluntary conditions before the switch (Figure 4.6), 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 4.3 and Figure 4.6). 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 4.5[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 seconds before the switch), confirming the main result of Experiment 1.

Figure 4.7 showed a significant difference between Delay/Voluntary and No Delay/Voluntary conditions 3-4 seconds 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 4.7 replicates the results of Experiment 1 and indicates once more that it is not the suspension

---

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

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 mid-task as in Experiment 1, which minimized the switching costs and made both conditions more similar. Behavioral results (Table 4.2) 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 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, i.e. thinking of things like "How many times more do I have to switch in order to have 3 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, since 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 3 times during a block and gave them a 30 second delay penalty at the end of the block if they switched less, which made the 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 4.3 and 4.6) 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. Since 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 seconds 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 seconds 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 seconds, a second music fragment played for 8 seconds plus 2 extra seconds 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 minutes, no matter how many blocks were completed. Participants were informed that the experiment would last 45 minutes 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 6 switches per memory game but it was never reached).

In order to see the effects of complete freedom in switching compared to 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 minutes of the Voluntary/No Delay condition of Experiment 1 and Experiment 2 and participants were again instructed to switch 3 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

25 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.

25 participants (9 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 (8 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 of Experiment 3

Behavioral Results

Two participants and 19 blocks from 6 other participants were removed from the music quiz version because they did not pass the

	Music quiz version	N-Back version
Average time per block (sec)	202.32 (11.87)	289.97 (18.14)
Average time per memory game (secondary task time removed)(sec)	171.92 (46.45)	197.36 (61.95)
Average nr of revisits per block	13.55 (1.51)	17.2 (2.2)
Average nr of lucky matches per block	1.21 (0.08)	1.26 (0.08)
Average nr of total clicks to complete a block	36.33 (1.5)	39.95 (2.23)
Average nr of switches to the secondary task per block	1.49 (0.18)	3.21 (0.09)
Resumption lag (sec)	1.93 (0.08)	1.91 (0.07)

Table 4.3. *Behavioral data of Experiment 3 (standard errors in the parenthesis)*

threshold of 3 seconds per card. Two participants and 10 blocks from 4 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 4.3 reports the behavioral data. In the music quiz version, participants spent on average 171.65 seconds per memory game. They completed on average 11.3 blocks (including in this analysis the 19 blocks that were removed from 6 participants but not the 2 participants that were completely removed) in approximately 45 minutes. 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 4 participants but not the 2 participants that were completely removed) in approximately 45 minutes. 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

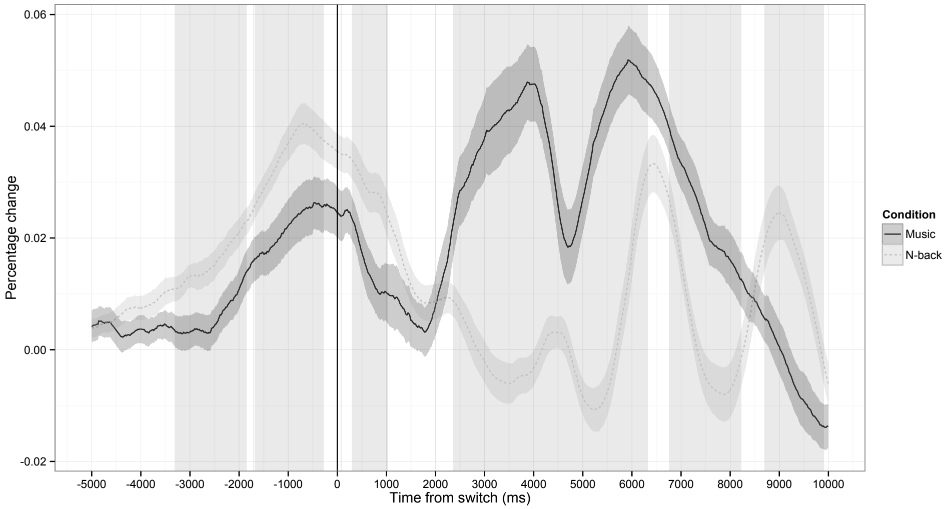


Figure 4.8. *Average pupil dilation for all participants of the 3rd experiment around the switch (0 is the switching point from memory game to the music quiz or the n-back task). Marked with dark grey color are the moments where there was a statistically significant difference (p<.05) between the music quiz version and the n-back version*

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, η_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, η_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=0.059, d=0.57). Resumption lag was not significantly different between the two conditions, with t(42.23)=0.19, p=0.85 and d=0.06.

Pupil dilation results

We performed t-tests between the music quiz and the n-back at every sample (10ms), with an FDR correction. The results are shown in Figure 4.8. 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, while the music quiz created a greater increase in pupil dilation than the n-back.

Discussion of Experiment 3

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 3 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, since the difference in time spent in the memory game was marginally significant in favour 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. Since 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 3 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 (Figure 4.8) indicates that this increase was not only due to mentally planning the switch points.

Figure 4.8 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 (Figure 4.5), 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 seconds 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 (Figure 4.3). 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 (Figure 4.6) 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 3 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 (Figure 4.8), 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, i.e. 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 4.8 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 2. This is a sign of rational multitasking behavior, since 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 (Table 4.2). 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, since 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 mid-subtask 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 introduction, 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-second 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 & Jones (2003). The delay also did not help for self-interruptions, since 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 seconds 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, since 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.

Applications

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, 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 et al., 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/

CHAPTER 6

Interrupt me!
External interruptions are less disruptive than self-interruptions.

This chapter was submitted (under revision) as:
Katidioti I., Borst, J.P., van Vugt, M.K., & Taatgen, N.A. Interrupt me! External interruptions are less disruptive than self-interruptions

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 and one in which pupil dilation was measured. 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 second, resulting in slower performance overall.

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 minutes (Gonzalez and Mark, 2004) and students interrupting their tasks every 6 minutes mostly to engage in social media (Rosen, Carrier and 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., Czerwinsky, Horvitz, and Wilhite, 2004; Gonzalez and Mark, 2004; Mark, Gonzalez, and 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 (Figure 5.1): 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

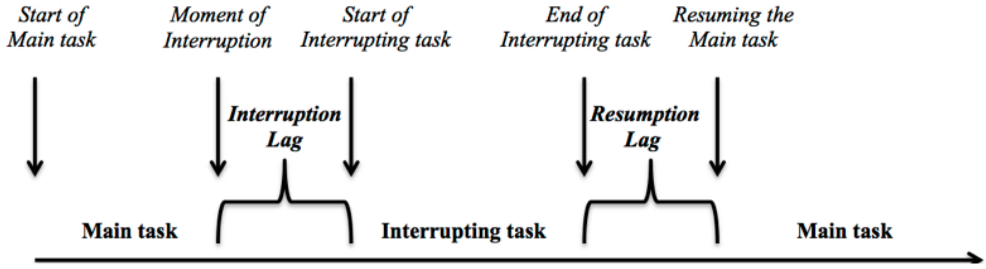


Figure 5.1. *Timeline of an interruption (based on Trafton, Altmann, Brock & Mintz, 2003)*

in pupil dilation before a self-interruption, which was not present before an external interruption (Katidioti, Borst and 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 out which kind of interruptions are less disruptive: self-interruptions or external interruptions.

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 minutes. 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 & Willie, 2000; Monk, Boehm-Davis, & Trafton, 2008; Trafton et al. 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 (Altman & 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 et al. (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 seconds 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. Czerwinsky, 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 et al. 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.

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, Gonzalez and Harris (2005), 52% of the interruptions in an office environment were self-interruptions and in the study of Czerwinsky, Horvitz, and Wilhite (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 & Benbunal-Fich, 2003). Other studies suggest that people self-interrupt or are distracted when the cognitive resources for the interrupting task are available. Taatgen, Katidioti, Borst & 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, & 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, 2015; 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 did compare external interruptions and 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 seconds) 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) and 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 minutes 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.

EXPERIMENT 1

Methods

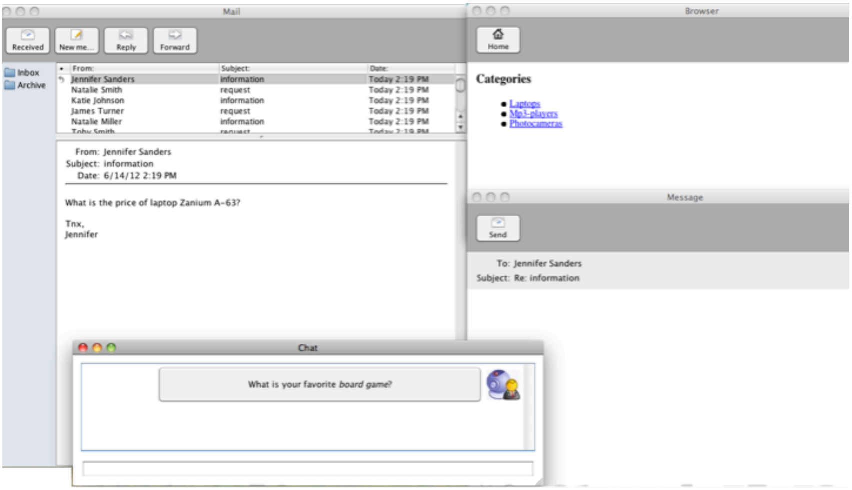
Design

The main and the interrupting task of this experiment were based on the ones used by Salvucci and Bogunovich (2010) and 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 Figure 5.2. 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 had to memorize the information in windows that were not currently visible.

The steps of the main task are shown in Figure 5.3. 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-second 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-second delay the price of the product loaded, the participant

---

[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.



From:	Subject:	Date:
Jennifer Sanders	information	Today 2:19 PM
Natalie Smith	request	Today 2:19 PM
Katie Johnson	information	Today 2:19 PM
James Turner	request	Today 2:19 PM
Natalie Miller	information	Today 2:19 PM
Toby Smith	request	Today 2:19 PM

From: Jennifer Sanders
Subject: information
Date: 6/14/12 2:19 PM

What is the price of laptop Zanium A-63?

Tnx,
Jennifer

Categories

- Laptops
- Mp3-players
- Photocameras

To: Jennifer Sanders
Subject: Re: information

What is your favorite *board game*?

Figure 5.2: *All the windows used in the experiment. During the experiment, the windows were overlapping.*

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 Figure 5.1, 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).

Conditions

Two factors were varied in the experiment: the kind of 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 (Figure 5.3) 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 seconds before deciding to self interrupt in the Voluntary block. In contrast, in the Forced block the interruption would occur immediately after the click to

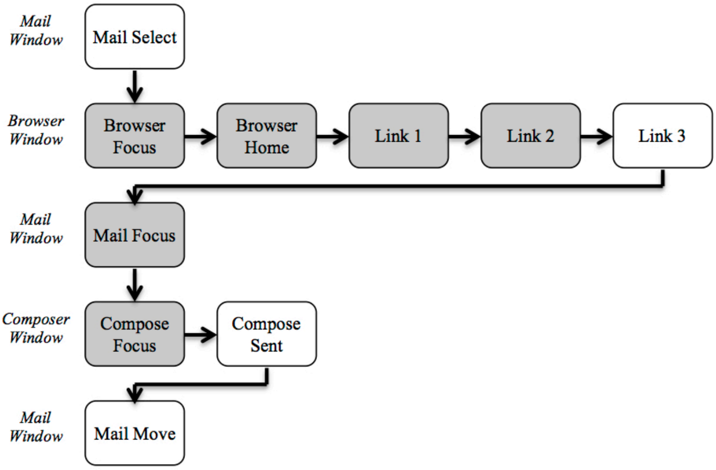


Figure 5.3. *The sequence of the main task. The high-workload moments are indicated with grey color and the low-workload moments with white.*

perform the main task move.

In the Delay blocks a 3-second 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-second loading time in Link 3 was present in both Delay and No Delay blocks. We used the delay manipulation in order to have longer high-workload moments. In Katidioti & 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.

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.

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.

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,

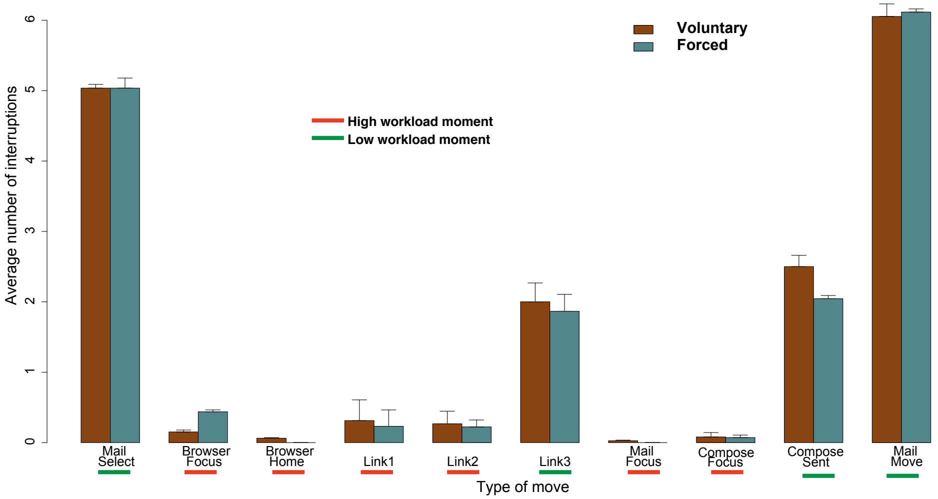


Figure 5.4. *Average number of interruptions per block for each step of the email task of Experiment 1.*

we only included the "Browser Focus" and "Browser Home" clicks that happened after "Mail Select" (as in Figure 5.2) 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.

Results

There were very few switches during the delay moments in the Delay condition – not nearly as many as in Katidioti & 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 (Figure 5.4). 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, Figure 5.4 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 seconds, SE=0.18), followed by the Forced condition (23.36 seconds, SE=1) and the Voluntary condition (24.83 seconds, SE=0.72). An ANOVA showed an effect of interruption type ($F(2,54)=27.03$, $p<.001$, $\eta_p^2= 0.5$) and follow-up t-tests confirmed that all conditions differed significantly from one another (all ps<.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 seconds (SE=0.063) and in the Forced condition 1.64 seconds (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-second 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 ($F(2,54)=1.45$, $p=0.24$, $\eta_p^2=0.05$).

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 (Figure 5.4), 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 seconds 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.

EXPERIMENT 2

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) and 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 and 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.

Method

Design.

The tasks used in this experiment were similar with Experiment 1.

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.

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 x 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 45cm 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.

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.

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.

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).

Results

Behavioral Results

Figure 5.5 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 ($t(20) = -19.17$, $p<.001$, $d=7.31$). 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 seconds, SE=0.57), followed by the Forced condition (21.78 seconds, SE=0.72) and finally

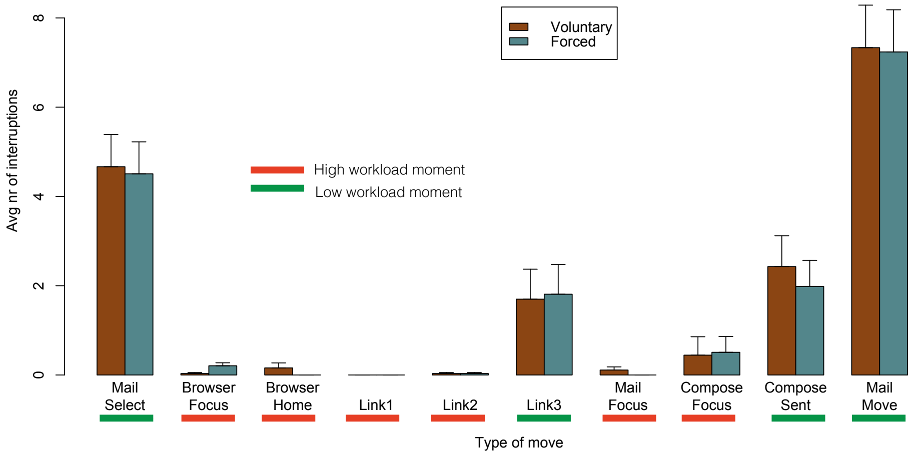


Figure 5.5. *Average number of interruptions per block for each step of the email task of Experiment 2.*

they were slowest in the Voluntary condition (22.71 seconds, SE=0.78). An ANOVA affirmed that this difference is significant (*F*(2,40)=8.48, *p*<.001, η_p^2= 0.3) and follow-up t-tests revealed that all conditions differed significantly from one another (*p*s<.05), Control and Forced marginally (*p*=0.07). The resumption lag was 1.49 seconds (SE=0.12) in the Voluntary condition and 1.53 seconds (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.

Pupil Dilation Results

Figure 5.6 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 seconds (SE=0.24) in the Voluntary condition and 3.66 seconds (SE=0.22) in the Forced condition. A t-test revealed that this

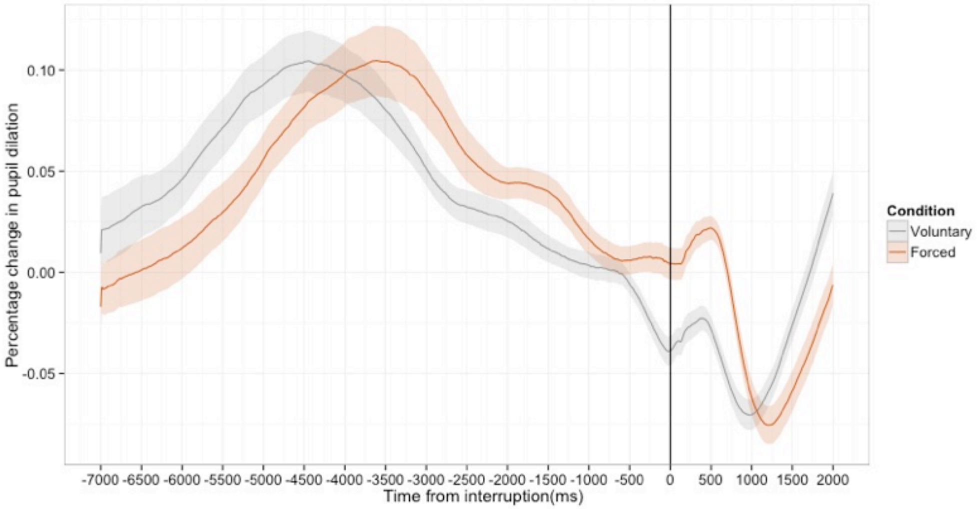
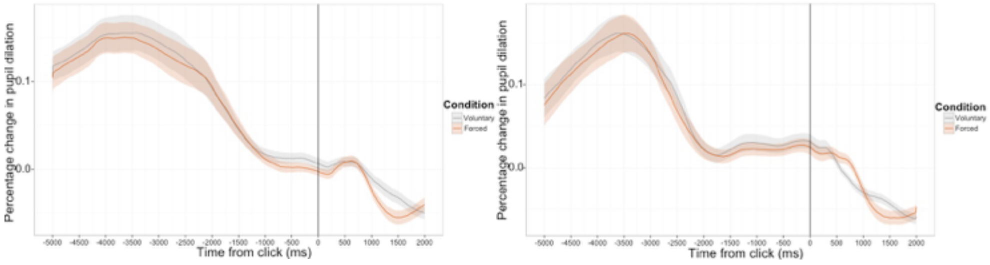


Figure 5.6. *Pupil dilation for the Voluntary and the Forced condition, time-locked to the moment of interruption. The vertical black line represents the moment of interruption.*



A

B

Figure 5.7. *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*

difference is significant ($t(20)=-4.17$, $p<.001$, $d=0.76$).

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 (Figure 5.7A) and the moment they drop an answered email in the Archive folder (Mail Move) (Figure 5.7B). Here, we did not observe any differences between conditions.

Returning to Figure 5.6, 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 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 ($t(20) = -1.72$, $p = 0.1$, $d=0.46$).

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 Figures 5.7A and 5.7B 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 (Figure 5.6), there is approximately a 1-second 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 second. 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 seconds 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.

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-interruptions (Figure 5.1). 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 (Figure 5.6). 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 5.8. 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 (Figure 5.4

and Figure 5.5). 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 and colleagues' (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.

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 and 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.

PART 3

Interruption Management System

CHAPTER 6

Interrupted by your pupil:
An interruption management system based on pupil dilation.

This chapter was submitted (accepted) as:
Katidioti I., Borst, J.P., Bierens de Haan, D.J., Pepping, T., van Vugt, M.K., & Taatgen, N.A. Interrupted by your pupil: an interruption management system based on pupil dilation. *International Journal of Human-Computer Interaction*

ABSTRACT

Interruptions are prevalent in everyday life and can be very disruptive. An important factor that affects the level of disruptiveness is the timing of the interruption: Interruptions at low-workload moments are known to be less disruptive than interruptions at high-workload moments. In this study we developed a task-independent interruption management system (IMS) that interrupts users at low-workload moments in order to minimize the disruptiveness of interruptions. The IMS identifies low-workload moments in real-time by measuring users' pupil dilation, which is a well-known indicator of workload. Using an experimental setup we showed that the IMS succeeded in finding the optimal moments for interruptions and marginally improved performance. Because our IMS is task-independent – it does not require a task analysis – it can be broadly applied.

INTRODUCTION

Nowadays it is nearly impossible for a work environment to be free of interruptions. Interruptions are often part of the job itself: it is hard to imagine a profession in which one focuses only on a single task for an extended amount of time. A pilot has to talk to the air traffic controller while operating a plane, a professor has to answer a student's question while giving a lecture, and a receptionist has to answer the phone in the middle of providing information to a visitor.

The prevalence of interruptions has been quantified by several observational studies that show how often people are interrupted in their workplace (e.g., Czerwinsky, Horvitz & Wilhite, 2004; Eyrolle & Cellier, 2000; Gonzalez & Mark, 2004). For example, Gonzalez and Mark's (2004) study revealed that office workers switched tasks every 3 minutes. In addition – and perhaps more worrisome – many other studies have shown that interruptions are very disruptive for the main task: users make more errors and become slower when interrupted (e.g. Edward & Gronlund, 1998; Gould, Brumby & Cox, 2013; Hodgetts & Jones, 2006; Iqbal & Bailey, 2007; Jin & Dabbish, 2009). As it seems impossible to ban interruptions from the workplace, it is crucial to find a way of managing interruptions that minimizes their negative effects.

With this goal in mind, we developed a task-independent interruption management system (IMS) that uses real-time pupil dilation measurements to interrupt users at the least disruptive moments of a task. We evaluated this system in a lab study, and showed that our IMS is able to find the optimal moments for interruptions. Before we describe our IMS in detail, we will first discuss what aspects of interruptions affect their disruptiveness and what kind of IMS's have been developed previously. We will then show how our IMS is able to identify the optimal interruption moments on the basis of pupil size independent of the current task.

Background

One of the main theories on interruptions is Memory for Goals (Altman and Trafton, 2002). In this theory, each task is characterized by a goal, which has a certain activation level. When a task is interrupted, its goal is stored in memory and starts decaying. In the meantime, the goal of the interrupting task is activated. Returning to the main task entails resumption of its goal. The longer the interruption lasts, the more the goal has decayed in declarative memory, and the harder it is to resume the main task.

In order to cover more of the factors that can affect the disruptive effect of interruptions, Borst, Taatgen and van Rijn (2015) extended Memory for Goals theory to Memory for Problem States. Instead of goals, this model focuses on problem states. The problem state contains the information that is necessary to complete the next steps in a task, e.g., when trying to find the value for x in an equation such as $2x + 4 = 14$, the information $2x - 10$ is stored in the problem state before proceeding to $10/2 = 5$. The main idea in Memory for Problem States is the same as in Memory for Goals: when the main task is interrupted, its problem state is stored and starts decaying. However, if the main or the interrupting task does not require a problem state, the main task will not be hard to resume even if a considerable amount of time has passed.

Memory for Problem States accounts for most of the factors that can affect the disruptive effect of an interruption. It is well known that interruptions disrupt the main task and affect performance (see for instance the seminal work by Gillie and Broadbent, 1989). However, there are multiple factors that affect the level of disruptiveness of interruptions. Several studies showed that a long interruption is more disruptive than a short one (e.g., Borst et al. 2015; Hodgetts & Jones, 2006; Monk, Boehm-Davis & Trafton, 2004), which is something that both Memory for Goals (Altman & Trafton, 2002) and Memory of Problem States account for, since the longer the goal or state of a task has to be stored, the harder it is to resume it. Other studies suggest that a more complex interruption is more disruptive than a simpler one (e.g., Borst et al. 2015; Cades, Boehm-Davis, Trafton & Monk, 2007; Monk et al., 2004). In addition, an interruption different from the main task is more disruptive than an interruption more similar to the main task (e.g., Gould et al., 2013; Iqbal & Bailey, 2008). The timing of the interruption during the main task also plays an important role. Interruptions at high-workload moments (typically in the middle of a (sub)task) are more disruptive than interruptions during low-workload moments (between (sub)tasks; Gould et al., 2013; Iqbal & Bailey, 2005; Iqbal & Bailey, 2006; Katidioti & Taatgen, 2014; Monk et al., 2004). In the current paper we will focus on minimizing the negative effects of interruptions by adjusting their timing.

In one of the studies focusing on the timing of interruptions, Gould and colleagues (2013) interrupted participants during a data-entry task either mid-subtask or between subtasks, with the former interruptions being more disruptive than the latter. Similar results were found by Monk and colleagues (2004) in a VCR programming task. Iqbal and Bailey (2006) used GOMS modeling to find high- and low-workload moments in three different tasks. They interrupted users at these moments, and showed that the cost of interruptions during low-workload moments was smaller than during high-workload moments.

In most of these studies, high workload was defined as participants' working memory being occupied and low workload as their working memory being unoccupied with task information. To test whether working memory requirements indeed play an important role in the disruptiveness of interruptions, Salvucci and Bogunovich (2010) created a real-life setup with clear high- and low-workload moments determined by working memory requirements (the current experiment is based on their study). In Salvucci and Bogunovich's study, participants simulated a client service employee for an electronics company, by answering emails from fictional clients asking them product prices. Participants had to read the email, look up the price in a browser, and write a response to the client. From time to time, a chat message arrived in the background. Participants were free to choose when to answer these messages. Results showed that participants preferred not to interrupt themselves during high-workload moments, which were the moments they had to remember the product name or the product price, causing their working memory to be occupied.

In a follow-up study by Katidioti and Taatgen (2014) that used the same email-and-chat task, participants were encouraged implicitly through an artificial delay in the main task to switch during high-workload moments in half the experiment. As a result, participants were slower to complete an email than when they switched at low-workload moments. Thus, interruptions at high-workload moments are more disruptive, and workload seems to be strongly dependent on working memory load.

Memory for Goals (Altman and Trafton, 2002) does not account for the effect the moment of interruption can have on the level of disruptiveness of an interruption, nor for the complexity of the tasks since the main focus of this theory is the effect of the length of the interruption. Memory for Problem States (Borst et al., 2015) can explain the effects of the moment of interruption on its disruptiveness (e.g., Gould et al., 2013; Iqbal & Bailey, 2005; Iqbal & Bailey, 2006; Katidioti & Taatgen, 2014; Monk et al., 2004). According to Memory for Problem States, if the main task does not require a problem state (and therefore the working memory is not occupied with task information), it will be resumed more easily after an interruption than a main task that requires a problem state, as no problem state has to be retrieved from memory in the former case.

Managing Interruptions

The fact that interruptions can be more or less disruptive based on the circumstances can be exploited by interruption management systems (IMS), which aim to find optimal – least disruptive – points for interruptions. McFarlane (2002) was one of the first to exploit this concept, and created an IMS that calculated the workload of the specific task he used and interrupted participants when the workload of the task was low. He then conducted an experiment in which he used a collection of performance and personal preference indices to compare four different kinds of interruptions: immediate, negotiated, mediated and scheduled. His results showed that mediated interruptions (determined by the IMS) were less damaging to performance than scheduled (occurring every 25 seconds) and immediate (random occurrence) interruptions. Negotiated interruptions (in which the user determined when to be interrupted) were comparable with the mediated interruptions on most indices, but required the user to make the decision when to switch. This suggested that a combination of these two systems would be beneficial for managing disruptions: a mediated system as the default, with the possibility to override the mediator and choose your own moment of interruption.

Arroyo and Selker (2011) developed an IMS that focused on the similarity between the interrupting task and the main task. Their IMS allowed relevant interruptions (defined as those with similar content as the main task) to pass, while it held back the irrelevant ones. This system led to a performance benefit for important/urgent tasks. Züger and Fritz (2015) used psycho-physiological measures (EEG data, eye blinks and electrodermal activity) to measure interruptibility of programmers. Although they did not create an IMS, they were able to identify a programmer’s state of interruptibility by means of machine-learning classifiers with high accuracy. This suggests that such classifiers could be potentially used in real time to interrupt users at low-workload moments. Tanaka, Abe, Aoki & Fujita (2015) and Kobayashi, Tanaka, Aoki & Fujita (2015) have developed an IMS that uses head motion and computer operations (typing, mouse clicks, opening and closing windows, etc.) in order to identify a user’s low-workload moments. Their system already shows very promising results, although it is limited to specific work environments and restricted to tasks that involve clicking, typing and window usage. This means, for example, that their system cannot find an interruption moment if one is reading a paper on a computer screen.

The goal of the current study is to find a way to interrupt people at low-workload moments. Therefore, we need a non-invasive way to measure workload. The studies reviewed above suggest that the best way to create a task-independent IMS is to use a physiological measure. The physiological measure we decided to focus on is pupil dilation.

Pupil Dilation

Since the 1960s, studies have established that the size of the pupil is not only affected by changes in light, but also by other stimuli. Hess and Polt (1960) were the first to show that pupil size also depends on covert cognitive variables. Nowadays, pupil dilation is known to react to a wide variety of cognitive processes such as task difficulty and working memory load (see Beatty & Lucero-Wagoner, 2000; Laeng, Sirois & Gredeback, 2011, for reviews). In general, it is clear that mental workload has a considerable impact on pupil size (e.g., Beatty, 1982; Hoeks & Levelt, 1993). A more difficult task evokes a larger pupil dilation than an easier task (e.g. Beatty & Lucero-Wagoner, 2000; Jennings & van der Molen, 2005). For instance, in a study by Kahneman, Tursk, Shapiro and Crider (1969), pupil dilation increased as the difficulty of mathematical equations increased.

There are many studies linking pupil dilation to working memory load – an important factor in mental workload. Kahneman and Beatty (1966) report that participants' pupil dilation increased as the number of digits they had to remember increased from 3 to 7. Pupil dilation increased again to the same size when participants had to repeat the digits for the second time. Peavler (1974) measured changes in pupil dilation while participants had to keep strings of 5, 9 and 13 digits in working memory. Results showed that pupil dilation kept increasing until the 7th or 8th digit and then reached an asymptote, reflecting the limits of working memory.

Besides working memory, pupil dilation is used in the study of many different forms of cognitive effort, operationalized as task complexity (e.g., Moresi, Adam, Rijcken, van Gerven, Kuipers, & Jolles, 2008; Prehn, Heekeren, & van der Meer, 2011), Stroop interference effects (Laeng, Ørbo, Holmlund, & Miozzo, 2011), or difficulty of retrieving information from memory (van Rijn, Dalenberg, Borst, & Sprenger, 2012). Despite the fact that pupil dilation is widely used in cognitive science, there is one drawback of using it as a real time measure: there is a delay of approximately one second before the pupil reaches its maximum dilation after the onset of a stimulus (e.g., Hoeks & Levelt, 1993).

Pupil dilation has also frequently been used in interruption research (e.g., Iqbal, Adamczyk, Zheng & Bailey, 2005; Katidioti, Borst & Taatgen, 2014). For example, Iqbal and colleagues (2005) found that pupil dilation decreased between subtasks, which are low-workload moments. In a follow-up study, Iqbal and Bailey (2005) found that there were smaller time costs when participants were interrupted at those low-workload moments, compared to high-workload or random moments. Combining all their findings, Iqbal and Bailey (2010) created the OASIS interruption management system, which delays interruptions until there is a natural breakpoint in the task. However, those breakpoints were decided by statistical models, based on behavioral data from previous studies (Iqbal & Bailey 2007; 2008) and not by real-time changes in pupil dilation. Thus, this system, as does the Arroyo and Selker (2011) IMS, requires a task analysis before it can be used.

The IMS we describe here chooses the optimal moments for interruption based only on changes in pupil dilation, independent of the specific task. When the user's pupil size drops below a certain threshold (which is constantly updated), it is considered a low-workload moment, and the user is interrupted. Since our IMS does not require a task analysis, it is task-independent.

INTERRUPTION MANAGEMENT SYSTEM

We developed an interruption management system (IMS) that uses real time changes in pupil dilation to identify the low-workload moments of a task. We then tested it on the email task that is interrupted by chat messages (Katidioti & Taatgen, 2014; Salvucci & Bogunovich, 2010) with minor adjustments to fit the current setup. As the task progresses, the IMS calculates a workload identifier value (WIV). When pupil size is below the WIV it is considered to be a low-workload moment. If there are consecutive low-workload moments for 200 ms[1], the IMS interrupts the participant (Figure 6.1). The WIV is calculated by using the following values: the baseline pupil size, the percentage change in pupil size (PCPS; Iqbal et al., 2005), the live average, and a threshold adapter.

The baseline pupil size is measured at the beginning of the study.

---

[1] The 200 ms interval was chosen after pilot studies. Because of blinks, saccades, and noise, a smaller interval might not have provided enough information. A bigger interval might have indicated wrong moments in the current task, which has quick changes from low to high-workload moments.

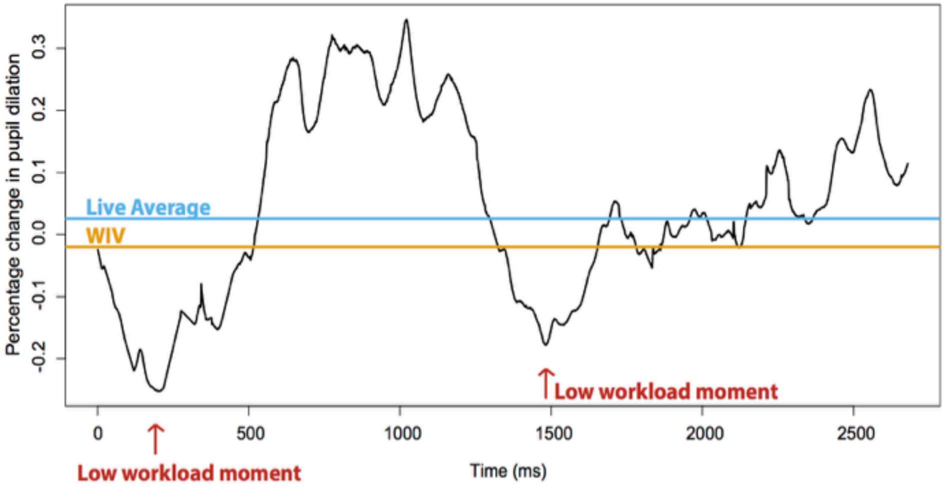


Figure 6.1. *Percentage change in pupil size (PCPS) during a random email sequence. The blue line is the value of the live average, the yellow line is the value of the WIV (workload identifier value) and the low workload moment of this email are indicated with red arrows. The IMS added 1000 to the PCPS values for calculation reasons, which is not shown in the figure.*

During the experiment, pupil size is measured continuously and then transformed into PCPS values by subtracting the baseline pupil size from each measurement and dividing the result by the baseline pupil size. In order to avoid multiplication with negative numbers, 1000 was added to each PCPS value. Thus, PCPS is given by Equation 1:

$$PCPS = \frac{current\ pupil\ dilation - baseline}{baseline} + 1000 \qquad (1)$$

The live average is defined as the average PCPS over the last minute and is used to account for possible changes in pupil dilation due to familiarity with the task, changing head position or changes in light (the latter two did not occur during the experiment).

The threshold adapter is set to 0.997 at the beginning of the task (the value was chosen after a pilot study). The threshold adapter is multiplied by the live average to calculate the WIV:

$$WIV = live\ average * threshold\ adapter \qquad (2)$$

The IMS allows for an interruption if pupil size measurements are below the WIV for 200 ms consecutively. In order to find the optimal WIV for each participant, the threshold adapter increases or decreases by 0.001 when there are more than one or no interruptions, respectively, during a specific time interval.

During the interruption, pupil measurements are not taken into account by the IMS, because the interruption is a task that may have different characteristics from the main task and pupil measurements may therefore not be representative. In addition, for 5 seconds after the end of an interruption, there cannot be another interruption. That restriction allows pupil measurement samples to return to baseline. Finally, eye blinks are ignored.

To test the IMS, we performed a lab study that tested whether the IMS could find the optimal interruption moments, and compared its performance to random interruptions and no interruptions at all.

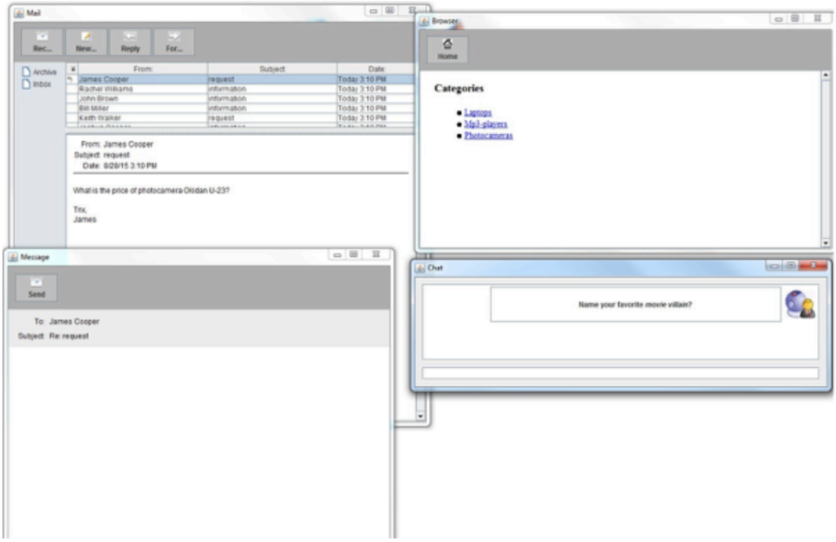
EXPERIMENT

Methods

The main and the interrupting task of the experiment were based on Salvucci and Bogunovich (2010). The experiment simulates the working environment of an employee of an electronics company who has to answer clients' emails while being interrupted by chat messages. The main task was the email-answering task and the interrupting task was the chat-answering task. The windows used in the experiment are shown in Figure 6.2. In the actual experiment, the windows were overlapping and could not be moved. The participant had to click on a window in order to see it. This forced them to remember the information in windows that were not currently visible.

The steps of the main task are shown in Figure 6.3. The participant first opens an email by clicking on it, reads the question (e.g. "What is the price of laptop Zanium A-63?"), goes to the simulated browser, clicks on the product category (Link 1), then the product name (Link 2) and finally the product code (Link 3). After a 2-second delay the price of the product loaded, the participant could read it, return to the email window and press the "Reply" button. The composer window would appear, the participant had to type the price, press the "Send" button (which caused the composer window to disappear) and then drag and drop the answered email in the Archive folder.

The interrupting task simulated a casual chat conversation. The



From:	Subject	Date
James Cooper	request	Today 3:10 PM
Rachel Williams	information	Today 3:10 PM
John Brown	information	Today 3:10 PM
Bill Miller	information	Today 3:10 PM
Keith Walker	request	Today 3:10 PM

From: James Cooper
Subject: request
Date: 8/28/15 3:10 PM

What is the price of photocamera Oklidan U-23?

Tnx,
James

Categories

- Laptops
- Mp3-players
- Photocameras

To: James Cooper
Subject: Re: request

Name your favorite movie villain?

Figure 6.2. *All the windows used in the experiment. In the top left corner there is the Mail window, top right corner is the Browser window, bottom left corner is the Composer window and bottom right corner the Chat window.*

chat questions were in the form of "What is your favorite...?" (e.g., color, food, movie, book). One in four questions was a follow-up question to the previous one, asking "Which is your least favorite?", which referred to the previous question. We used these follow-up questions to engage the participants more into the simulated conversation. When an interruption occurred, the chat window appeared in front of the other windows and could not be unfocused until the participant responded. Participants were instructed to immediately answer one chat message and then continue with the email task. Although in our experiment the chat questions were of a private nature, this simulates situations in some working environments in which the employees have to give priority to the live-chat questions that clients ask them.

The email task has high- and low-workload moments (see Figure 6.3). High workload moments are considered the moments where working memory was occupied by either the product name or the product price. Low-workload moments are defined as the moments that working memory was not occupied by task information. At first, after opening the email, the participant has to memorize the product type and name until finishing the search (Link 3). Link 3 is a low workload moment, since participants

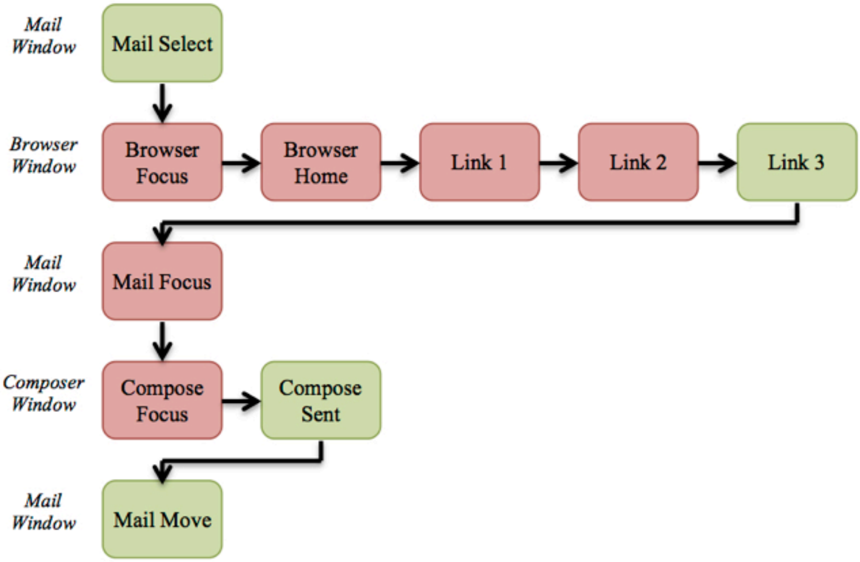


Figure 6.3. *The sequence of the main task. The high-workload moments are indicated with red and the low-workload moments with green.*

no longer have to retain the product name in working memory. When the price of the product loads, there are again high workload moments, since participants have to keep the price in their working memory until the answer is sent. A similar task analysis has been used at Salvucci & Bogunovich (2010) and Katidioti & Taatgen (2014). Both these papers have used the same e-mail task (with some small differences in Salvucci and Bogunovich, 2010). Both these papers gave participants the freedom to self-interrupt and results confirm this task analysis of high and low-workload moments.

Conditions

We used a within-subject repeated-measurement design with three conditions: Control, IMS and Random. During a Control block there were no interruptions: the participant only had to perform the email task to measure baseline performance. In a Random block, interruptions occurred at random moments. At the beginning of the block a random interruption moment between 10 and 30 seconds in the future was picked. The pilot study showed that the average time to complete an email sequence is about 20 seconds, so this interval would result in approximately one interruption per e-mail. When the designated moment arrived in the experiment, an interruption occurred. After the interruption finished, a new interruption moment between the next 10 and 30 seconds was picked for the next interruption, etc.

In the IMS blocks, the moments of the interruptions were determined by the IMS. As in the Random condition, a random moment between 10 seconds and 30 seconds was picked. The time until that moment was the interruption interval. If the IMS detected no suitable interruption moments during the interruption interval, the threshold adapter was increased by 0.001, so that an interruption would be more likely in the next interval. If the IMS interrupted the participant more than once during the interruption interval, the threshold adapter was decreased by 0.001 each time an extra interruption happened. The time spent on the interruptions was added to the interruption interval. When the interval finished, another random moment between 10 and 30 seconds was picked, etc.

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 x 1200 pixels and screen density of 64 pixels/inch. Participants were asked to use a chin-rest during the experiment. The eyetracker was an Eyelink 1000 from SR Research, positioned approximately 45cm from the end of the desk.

Eye fixations were 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. The eye tracker's default parameters were used to convert gaze positions into fixations and saccades.

Procedure

Participants started with a practice phase of 6 uninterrupted emails, during which the baseline pupil dilation was calculated. After the sixth email was archived, the first block started immediately.

The experiment consisted of three parts, each of the parts containing three blocks in random order: one Control block, one Random block and one IMS block. Each block finished after 10 emails were archived and the participant could then take a break. The experiment finished after all 9 blocks were completed. The experiment lasted approximately 50 minutes.

Participants

26 (19 female) participants were tested. Four participants were removed because they had at least 2 blocks where the IMS hardly interrupted them. The reason for the scarce interruptions is that it took a long time before the IMS managed to find the optimal WIV for these participants. A possible explanation for that is that the original threshold value (0.997) was too high for these participants and the IMS kept interrupting them in the beginning of the block. Although the IMS lowered the threshold, the fact that they had to deal with so many interruptions and typing of the answers probably made the participants' pupils dilate, resulting in a threshold that was too low for them. If the blocks had been longer, the IMS might have managed to find the optimal threshold. We decided to remove these participants because their uninterrupted IMS blocks are not representative of how the IMS works.

The remaining 22 participants (15 female) had a mean age of 23.2. All participants had normal or corrected-to-normal vision, gave informed consent for their participation and received monetary compensation of 8 euros.

Preprocessing

Two blocks were removed because participants did not follow instructions, one from the control condition and one from the IMS condition. Furthermore, 62 emails were removed because each of them had 4 or more interruptions. Only 6 of these belonged in the random condition, the rest were mostly the first emails of the IMS condition, when the IMS needed some time to find the appropriate WIV value.

In order to construct Figure 6.7 – not for the IMS – eye blinks were removed from the pupil dilation data, starting 100 ms before the blink up to 100 ms after the blink. The removed pupil dilation data were replaced by a linear interpolation and then all data were down-sampled to 100 Hz. We calculated the percentage change in pupil dilation from a baseline, which was defined by a very slow lowess filter[2] (a smooth curve that follows pupil dilation, with the parameter values as provided by R, 2008) given by a weighted linear least squares regression over the span (following Katidioti et al., 2014).

---

[2] Parameters used: smoother span = 2/3, number of robustifying iterations that should be performed = 3, delta=0.01 * the range of time

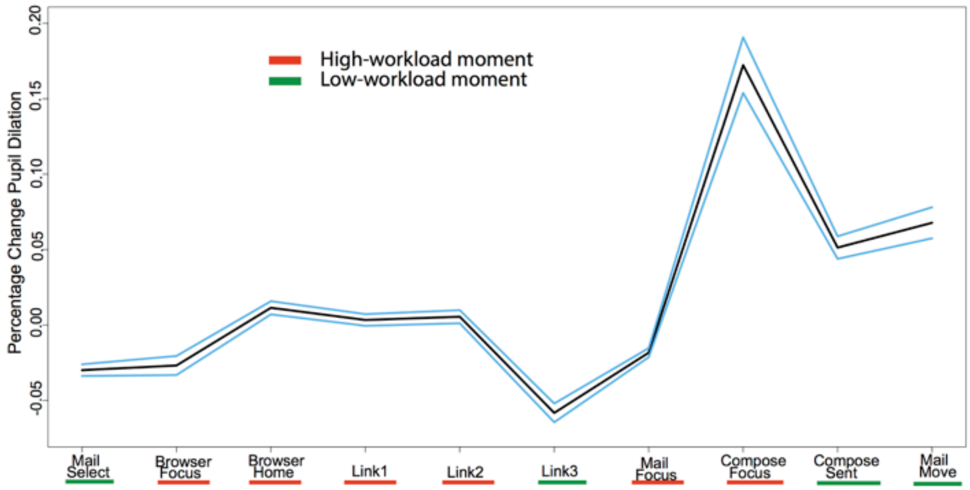


Figure 6.4. *Average PCPS values for every move for uninterrupted emails (Control condition). The blue lines show the standard error.*

RESULTS

Pupil Dilation and Interruption Moments

Figure 6.4 shows the average percentage change in pupil dilation (PCPS) values at each email sequence moment for the Control condition, that is, emails that contained no interruptions. It is clear that pupil size increases on high workload moments and then decreases on low-workload moments. The IMS identifies the low-workload moments by the comparing the percentage change in pupil dilation of the last 200 ms with the threshold. Overall we observe the highest PCPS during the "Compose Focus" move, probably because participants had to type at that point. That is the reason why—although there is a large drop in the PCPS in the next step ("Compose Sent")—the value is still higher than that of other high-workload moments.

Link 3 is an interesting point in the task. It is a low-workload moment but only lasts about 2 seconds and it is in the middle of the task, between high-workload moments. Nevertheless, the size of the pupil decreases at this point (Figure 4) and the IMS is able to detect that decrease (see Figure 6.5).

Figure 6.5 shows the number of interruptions for each moment in the email sequence. The IMS succeeded partly in shifting the interruptions

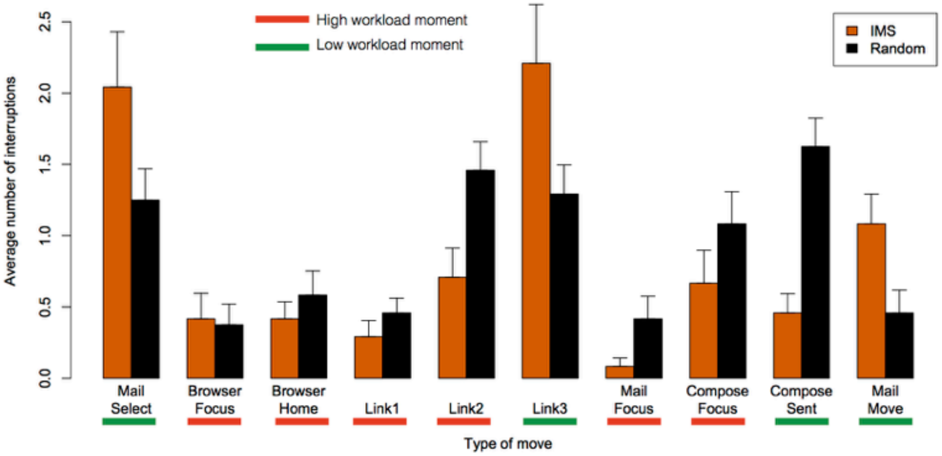


Figure 6.5. *Average number of interruptions that occurred in each step of the main task for the IMS and the Random conditions. The high workload moments are indicated with a red line and the low workload moments with a green line.*

to the low-workload moments (green moments in Figure 5). Compared to the Random condition, using the IMS resulted in more low-workload moment interruptions and fewer high-workload moment interruptions. On average, there were 10.27 interruptions per block in the Random condition (4.7 during high-workload moments and 5.58 during low-workload moments) and 6.26 interruptions per block in the IMS condition (2.15 during high-workload moments and 4.1 during low-workload moments). Proportionally this means that in the IMS condition 66.8% of the switches occurred at low-workload moments and 33.2% at high-workload moments, whereas in the Random condition the percentages were 54.0% and 46.0% respectively (Figure 6.6).

According to a two-way repeated measures ANOVA with Condition (IMS vs Random) and Workload (low vs high) as factors, there were significantly more low-workload than high-workload interruptions ($F(1,21)=21.15$, $p<.001$, $\eta_p^2= 0.5$) and significantly more interruptions in the Random condition than in the IMS condition ($F(1,21)=23.89$, $p<.001$, $\eta_p^2= 0.53$). Most importantly, the interaction between Condition and Workload was significant ($F(1,21)=5.29$, $p=.032$, $\eta_p^2= 0.2$), indicating that the IMS changed the proportion of interruptions to more interruptions at low-

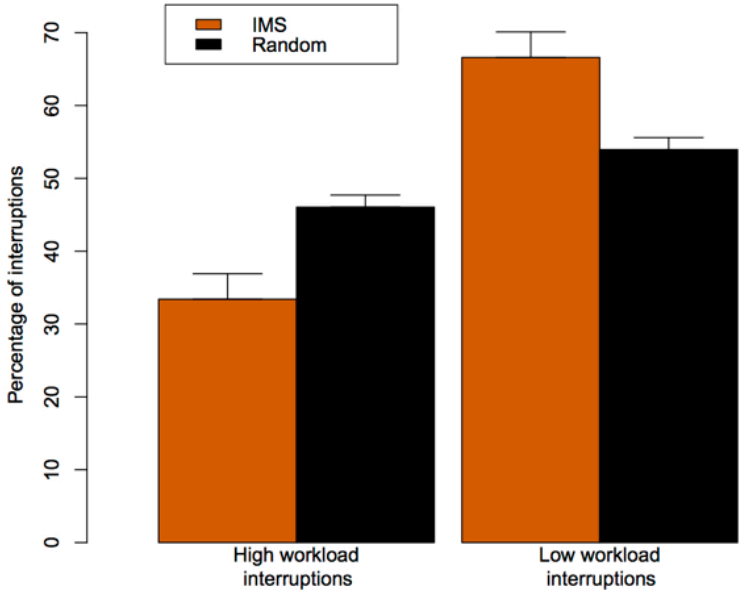


Figure 6.6. *Average number of interruptions on high and low workload moments per block for both conditions*

workload than at high-workload moments. Furthermore, block analysis showed that the IMS increased the percentage of low-workload interruptions as the blocks progressed, from 58.04% in the first IMS block to 67.38% in the last IMS block.

Taking into account the one-second delay in pupil dilation reaction (e.g., Hoeks & Levelt, 1993), we checked what type of moment occurred 1.1 – 0.9 seconds before the high-workload moment interruptions that the IMS created. Results showed that at 62.2% of the time that was indeed a high-workload moment and 37.8% of the time it was a low-workload moment.

To assess whether the IMS interrupted users when their pupil dilation was low, as we intended, we compared the average pupil dilation for the Random and the IMS condition around the interruption point, time-locked at the moment of interruption (time = 0 sec, Figure 6.7). This figure shows that the IMS indeed interrupted users when their pupil dilation was low. In both conditions there is an increase in pupil dilation approximately 700 ms after the interruption. This increase likely reflects the pupil's reaction to the interruption itself.

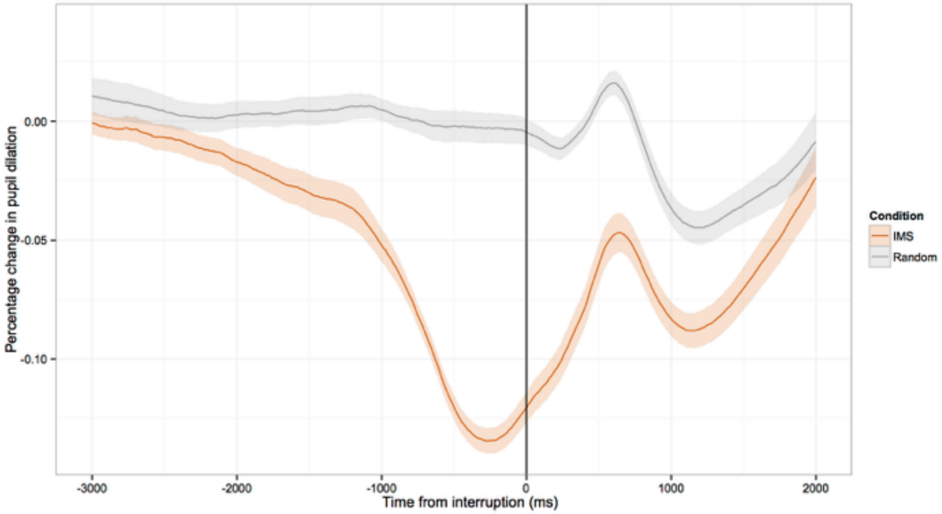


Figure 6.7. *Average pupil dilation around the interruption point (indicated by the black line) for the IMS and the Random conditions. The lighter area around each line represents the standard error.*

Performance

In order to verify that high-workload moment interruptions are worse than low-workload ones, we compared emails in which participants were interrupted at a high-workload moment to emails in which they were interrupted at a low-workload moment (independent of condition). When interrupted at a high-workload moment, participants were significantly slower in completing the email sequence (22.37 seconds) than when they were interrupted at a low-workload moment (20.21 seconds; t(21)=-3.76, p=.0012).

The success of the IMS in detecting low-workload moments was reflected in the participants' performance on the main email task. The average time to complete an email (after removing the time spent on interruptions and emails that deviated more than 3 SDs from the mean) per condition was 18.16 (SE=0.73) seconds for the Control condition, 20.30 (SE=0.71) seconds for the IMS condition and 21.53 (SE=1.19) seconds for the Random condition. An ANOVA revealed a significant difference between conditions (F(2,42)=19.18, p<.001, η_p^2=0.48), and a pairwise t-test with Bonferroni-Holm correction revealed that the difference between the IMS and the Random condition was marginally significant (t(21)=-2.05, p=.053). Participants were significantly faster in the Control condition then in the IMS or the Random condition (both ps<.001).

Participants seldom made any mistakes (such as typing the wrong price or looking up the wrong product). However, there were times that participants forgot the product name while browsing and revisited the email window in order to read it again. There were on average 0.32 (SE=0.05) revisits per email in the Control condition, 0.44 (SE=0.08) in the IMS condition and 0.46 (SE=0.07) in the Random condition. An ANOVA showed that this difference is significant (F(2,42)=4.567, p=.016, η_p^2=0.18), and a follow-up pairwise t-test showed that the only significant difference was between the Control and the Random condition (t(21)=-2.82, after a Bonferroni-Holm correction p=.031).

Another performance measure used often in interruption studies is the resumption lag: the time one needs to resume the main task after being interrupted. In this setup, resumption lag is the time between sending the answer in the chat and the next main task move. The resumption lag was 1.55 seconds for the IMS condition and 1.37 seconds for the Random condition, a difference that was not significant according to a t-test (t(21) = 1.37, p=0.18, d=0.37).

DISCUSSION

The aim of this study was to create and test an interruption management system (IMS) that interrupts users at optimal moments of an ongoing task. It has been shown by many studies (e.g., Gould et al., 2013; Iqbal et al., 2005; Iqbal & Bailey, 2005; Katidioti & Taatgen, 2014; Salvucci & Bogunovich, 2010) that interruptions at low-workload moments are less disruptive than interruptions at high workload moments. In order to measure workload and find low-workload moments, we used pupil dilation, a well-known measure of cognitive workload (e.g. Beatty, 1982; Beatty & Lucero-Wagoner, 2000; Hoeks & Levelt; 1993; Iqbal et al. 2005). We developed a task-independent IMS that interrupts users when their pupil dilation drops below an adaptive value (Figure 1). We then tested this IMS in an experimental study, using an email and chat setup that resembles client service in an electronics company (Salvucci & Bogunovich, 2010).

PCPS increased on the known high-workload moments of this task and decreased on the low-workload moments (Figure 4), confirming that pupil dilation reacted to the workload changes. Behavioral results showed that the IMS succeeded in interrupting participants on the low workload moments of the main task (Figure 5). Even Link 3, which is a low-workload point lasting only about 2 seconds was detected by the

IMS, as seen in Figure 6. In the Random condition, switches were almost equally distributed between low workload (53.95%) and high workload (46.05%) moments. In the IMS condition, the IMS managed to increase the difference, by interrupting participants 66.81% of the time on low workload moments. In addition, pupil dilation results confirmed that these interruptions happened when pupil dilation was low or decreasing in the IMS condition (Figure 7). Taking all of the above into account, we can conclude that the IMS was successful in interrupting people at low-workload moments by detecting a decrease in their pupil dilation.

Performance in this setup was measured by the average time needed to complete an email sequence. High workload interruptions (across both conditions) made participants significantly slower than low-workload interruptions. This result is in line with previous studies on the timing of interruptions (Gould et al., 2013; Iqbal et al.,2005; Iqbal & Bailey, 2005; Katidioti & Taatgen, 2014; Salvucci & Bogunovich, 2010), which all suggest that low-workload interruptions are less disruptive than high-workload interruptions.

Although there was no difference between being interrupted by the IMS or randomly in some performance measures (resumption lag and number of revisits to the email window), participants were marginally faster to complete an email in the IMS than in the Random condition. This is promising, especially given that the IMS did not result in 100% low-workload switches, but only 67%. There are two possible explanations for that. The first explanation is that since pupil dilation reacts to stimuli with a one-second delay (Hoeks & Levelt, 1993), the behavioral task we used alternated too quickly from low- to high-workload moments for the IMS to perform optimally – and consequently, 100% low-workload switches were impossible. By the time the IMS decided that the pupil dilation was low enough to interrupt, the low-workload moment of the task might have already changed to a high-workload one – for example opening an email is a low-workload moment, but reading it lasts only a couple of seconds and then working memory is occupied again. Analysis of the task moments that occurred 1.1 – 0.9 seconds before high-workload interruptions during the IMS blocks revealed that they were mostly also high-workload moments (62.2%). Thus, the remaining 37% of the high-workload switches might be due to a low-workload moment about a second earlier. The second explanation is that the IMS needs some time to find the optimal WIV. In some cases, that led to participants being constantly interrupted in the beginning of the IMS blocks, or not interrupted at all. The fact that the percentage of low-workload interruptions increased from 58.04% in the first IMS block to 67.38% in the last IMS block supports this explanation. In a real-life environment this should be less of a problem if pupil dilation can be measured continuously and the WIV updated throughout the day.

A substantial advantage of our IMS is that it is task-independent. Some parameters (i.e., the 10-30 second interval between interruptions that was used in order to have a fair comparison with the Random condition, the 200ms pupil dilation judging period and the rate/amount of the threshold adaptor changing) were chosen to fit this specific task, such that we could make a valid comparison between the IMS and random interruptions. However with few minor changes (e.g., having only a set number of interruptions per hour), the IMS can be adapted to different tasks. The IMS is task-independent because it interrupts people only based on the changes in their pupil dilation, not on the properties of the main task. It can therefore be applied without first having to perform a task analysis, which is required for most other systems (e.g., Arroyo & Selker, 2011; Iqbal & Bailey, 2010). Furthermore, the pupil dilation during the interruption is not taken into account, which means that the IMS is not affected by the extent to which the interrupting task is relevant to the main task. Since our IMS is task-independent, it is possible to integrate it into an operating system and use it across tasks. For instance, simple office work could benefit from such an IMS, which could defer email and social media notifications until a low-workload moment. Naturally, it could also be employed in single-task environments such as the cockpit or air-traffic control. In those cases, the current system might be combined with a task analysis to identify crucial processes that may never be interrupted.

One issue with a pupil-dilation-based IMS is that one needs to measure pupil dilation in real time in an uncontrolled environment, where lighting conditions might affect the pupil. Although pupil dilation has traditionally been measured with expensive eye-tracking systems, in recent years webcams have rapidly become more capable, to the extent that most current webcams are high definition. Such high-quality webcams can be used to measure pupil dilation changes and calculate workload in normal office conditions (Rafiqi, Wangwiwattana, Fernandez, Nair, & Larson, 2015), promising to make eye tracking and pupil dilation widely available.

The IMS used a specific algorithm that compares 200 ms worth of pupil dilation to an adaptive threshold in order to judge whether the pupil dilation indicated a low-workload moment. We could change the algorithm to fit the task better, by comparing the pupil dilation of each step of the task to the pupil dilation of the previous step. This would be a better idea for this task, since the changes from a high to a low-workload moment can be very quick. Although this change would probably yield better results with this specific task, it would also make the IMS task-dependent. Another idea is to take the direction of the pupil change into account. Only if the pupil size keeps decreasing for a specific amount of time, then the workload decreases and it is a good moment for an interruption. Although this seems a good idea on the basis of the average data, its robustness still has to be investigated online.

Interruption management systems (IMSs) are becoming a necessity in the world we live in that is full with interruptions that endanger our work. There are many studies that point out what makes interruptions disruptive and how to minimize their negative effects (e.g., Edward & Gronlund, 1998; Gould, Brumby & Cox, 2013; Hodgetts & Jones, 2006; Iqbal & Bailey, 2007; Jin & Dabbish, 2009) but few have implemented this knowledge in a usable system for managing interruptions. In the study presented here we focused only on one aspect of interruptions (timing of the interruption) and one psycho-physiological measure (pupil dilation), avoiding the need for extensive task analysis that is required in many previous systems. We showed that a pupil-dilation-based IMS can identify low-workload moments in real time, and interrupt users at opportune moments, leading to marginally better performance than random interruptions. Although our IMS can be further optimized – for instance by taking into account other sources of data (Züger & Fritz, 2015) – it already showed promising results.

CHAPTER 7

SUMMARY AND CONCLUSIONS

This thesis focused on self-interruptions: the constant need we all have to stop our work and engage in other activities, such as checking social media, having coffee breaks or staring out of the window. Self-interruption is a very important issue that seriously affects important aspects of everyday life, such as working or studying (e.g. Mark, Gonzalez, & Harris, 2005; Rosen, Carrier, & Cheever, 2013). Although interruptions have received a great deal of attention by the scientific community, experimental studies mainly focus on external interruptions, since self-interruptions are harder to evoke in an experimental setup.

In this thesis we investigated what can create self-interruptions, compared them to external interruptions and created a system that tried to eliminate the need for self-interruptions by creating less disruptive external interruptions. In this final chapter I provide a summary of the findings.

Self-interruption and cognitive resources

Previous studies have shown that people self-interrupt rationally (e.g., Salvucci & Bogunovich, 2010) in the sense that they avoid self-interruptions during high-workload moments, which is known to be more disruptive than switching during low-workload moments (e.g. Monk, Boehm-Davis & Trafton, 2004). In Salvucci and Bogunivich's (2010) study (and all the interruption studies mentioned in this thesis) a high-workload moment is a moment within the task during which focal working memory (see Oberauer, 2002) is occupied with some piece of information (e.g. with a product model or a product price, in the case of Salvucci and Bogunovich, 2010).

In Chapter 2 we challenged this rationality by creating a situation where cognitive resources other than working memory (vision and motor functions, which were required for the interrupting task) were free during high-workload moments. Participants self-interrupted during those moments, which affected their main task performance negatively. We concluded that when the cognitive resources for another task become available, people will self-interrupt even when this leads to worse performance. This result is in line with the resource usage principle of Threaded Cognition Theory (Salvucci & Taatgen, 2011), which states that if two tasks are competing for the use of a resource, once this resource becomes available from one task, the other task will immediately use it.

In Chapter 3 we delved deeper into the association between cognitive resource availability and self-interruption, or to be more precise, distraction. We used two tasks that required the use of different cognitive resources, a visual task and a problem-solving task. Both tasks had three difficulty levels (easy, medium, hard), with the visual task requiring more visual resources as it became harder and the problem-solving task requiring fewer visual resources as it became harder. There was also a visual distractor (a silent video) present in the periphery of participants' visual field. We used an eye-tracker to monitor the distractibility of participants by the video. Results showed opposite patterns for the two tasks: as the visual task became more difficult, participants were less distracted by the video, whereas as the problem-solving task became more difficult, they were more distracted by the video. The difficulty of the task did not affect distractibility; the availability of visual resources did.

Our results in Chapter 2 and Chapter 3 are in line with Threaded Cognition Theory (Salvucci & Taatgen, 2011) and Multiple Resources Theory (Wickens, 2002). Both theories agree that tasks sharing a cognitive resource (e.g., two visual tasks) are hard to combine. Our results change the vantage point from task to distraction: when the cognitive resources for an interrupting task become available, people will be distracted by this task and self-interrupt. These results demonstrate how difficult it is to block self-interruptions. Even if we are locked in an empty room in order to focus on work, when mental resources become available, we will daydream.

External vs. self-interruptions

If asked, people would intuitively say that a self-interruption is less disruptive than an external interruption. Choosing when to interrupt yourself feels better than being unexpectedly interrupted. However, there is little experimental evidence to support this intuition. Most of the few studies comparing self-interruptions with external interruptions (e.g., McFarlane, 2002; Panepinto, 2010) have some design weaknesses that render their results difficult to interpret.

In this thesis, we compared external interruptions with self-interruptions in two different studies. In Experiment 1 of Chapter 4 we showed that there was no difference in performance in favor of the self-interruption. This is significant because external interruptions occurred at high-workload moments, while participants naturally chose to self-interrupt at low-workload moments. This suggests that in fact external interruptions would have been better for task performance if the workload at the

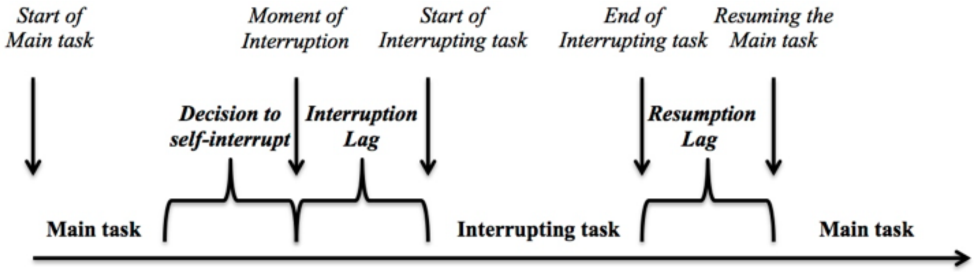


Figure 7.1. Self-interruption timeline

moment of interruption had been the same. In Experiment 2 of Chapter 4 we demonstrated this directly, by matching external interruptions to a participant's self-interruptions from the previous block. Results showed that in that setup, external interruptions were less disruptive than self-interruptions.

We replicated that result in Chapter 5 using a different task. Once more, we compared self-interruptions with external interruptions occurring at similar moments. The results agreed with those of Chapter 4: external interruption was less disruptive than self-interruptions.

It is important to note that we do not suggest that all kinds of external interruptions are less disruptive than self-interruptions (as is obvious from Experiment 1 of Chapter 4). In fact, we showed that only low-workload external interruptions are less disruptive than self-interruptions. We cannot draw conclusions about external interruptions during high-workload moments, since we only used this setup in one experiment (comparing them to self-interruptions which occur as usual at low-workload moments) and found no difference in their disruptiveness.

Pupil dilation and self-interruption

In Chapters 4 and 5, in addition to the behavioral results, we compared the changes that occur in pupil dilation during the interruption moment for external and self-interruptions. Results showed that there was a difference between the two kinds of interruption, with self-interruption creating a larger increase in pupil size time-locked to the interruption. The pupil dilation continued to increase during the self-interruptions for a longer period than it did for the external interruptions.

Both pupil dilation and behavioral results of our experiments suggest that there is a difference in the interruption timeline between external and self-interruptions. This implies that the interruption timeline proposed by Trafton et al. (2003) (Figure 1.1 in Introduction of this thesis) is not complete when it comes to self-interruptions.

In Chapter 4 we considered and rejected a number of different reasons that could explain the difference in pupil dilation between external and self-interruptions. We concluded that the decision to self-interrupt is the only explanation behind the greater reaction of pupil dilation during the moment of a self-interruption. We used this information to update the interruption timeline for self-interruptions as seen in Figure 7.1.

Interruption management system

Since our experimental results showed that low-workload external interruptions are less disruptive than self-interruptions, we wanted to apply this finding in a real-world context. We decided to create an interruption management system that would locate low-workload moments within a task and interrupt people at those moments, making the need for a self-interruption obsolete (Chapter 6). We used changes in pupil dilation as a tool to locate the low-workload moments of the task.

It is known that pupil dilation increases as cognitive workload increases and vice versa (e.g. Beatty & Lucero-Wagoner, 2000). Therefore, we created an algorithm that used real-time changes of pupil dilation to choose the optimal moment for an external interruption. We used that system in an experimental study and compared system-generated interruptions to random interruptions. Results showed that our interruption management system increased the percentage of interruptions during low-workload moments and marginally improved task performance. These results are promising. We think that with some small tweaks, this interruption management system could be a very useful tool for various working environments.

Conclusion

In this thesis, interruptions are viewed as an unavoidable part of everyday life. As hard as we try, it is impossible to have an interruption-free working day, especially a self-interruption-free working day. The results of our experiments offered new insights on when self-interruptions occur and how we could make them less disruptive.

The main findings of this thesis are that:

- If the cognitive resources needed by an interrupting task are free, the possibility for a self-interruption increases
- The decision period before a self-interruption has time costs that

are not present in an external interruption. Therefore an external interruption during a low-workload moment tends to be less disruptive than a self-interruption.

These results can be applied in everyday life to reduce the amount of self-interruptions and minimize their disruptive effects. Avoiding distractors in the workplace that do not share cognitive resources with our main task and the use of interruption management systems (like the one we implemented) can create a big difference in a working or studying environment.

Samenvatting

Deze samenvatting van 852 woorden gaat over een onderwerp dat ik goed ken. Desondanks kostte het me meer dan een dag om te schrijven. De reden daarvoor is dat mijn schrijven vaak werd onderbroken, zowel door externe factoren als door zelf-onderbrekingen. Over de externe onderbrekingen had ik meestal geen controle: meetings, pratende collega's, rinkelende telefoons… Maar zelfs als ik probeerde om deze onderbrekingen te vermijden door mezelf af te zonderen nam ik de tijd om mezelf te onderbreken om mijn emails nog eens te controleren of om uit het raam te staren. Dit zal veel mensen vast bekend voorkomen. Onderbrekingen zijn een alledaags probleem dat onze productiviteit beïnvloedt. Vooral zelf-onderbrekingen zijn moeilijk te begrijpen en te vermijden. In dit proefschrift heb ik geprobeerd om meer te leren over zelf-onderbrekingen en om manieren te vinden om hun storende werking te verminderen.

De invloed van beschikbare cognitieve middelen op zelf-onderbrekingen

Wat gebeurt er wanneer je een chatbericht krijgt tijdens een druk moment in het uitvoeren van een taak? De participanten in ons experiment in Hoofdstuk 2 vertoonden rationeel gedrag: ze wachtten totdat ze geen informatie meer hoefden te onthouden voordat ze antwoord gaven op het chatbericht. Maar wanneer de browser voor oponthoud zorgde stopte het rationele gedrag: ze kozen ervoor om het chatbericht te beantwoorden tijdens het oponthoud, vergaten de informatie die ze moesten onthouden en verloren daardoor meer tijd.

Wordt je sneller afgeleid door een kattenvideo in de achtergrond wanneer je een simpele taak uitvoert, of juist een moeilijke? Wanneer participanten in ons experiment in Hoofdstuk 3 een visuele taak uitvoerden, keken ze minder vaak naar de video wanneer de taak moeilijker werd, omdat hun visuele systeem meer werden bezet door de taak. Wanneer ze echter een taak uitvoerden waar ze een probleem moesten oplossen, waren ze meer afgeleid door de video naarmate de taak moeilijker werd, omdat hun visuele systeem minder bezet werd. Deze resultaten laten zien dat de moeilijkheidsgraad van de taak niet belangrijk was – de beschikbaarheid van cognitieve middelen was wat participanten meer of minder afgeleid maakte.

In het eerste deel van mijn proefschrift heb ik gevonden dat mensen zichzelf vaker onderbreken wanneer de middelen die nodig zijn voor de zelf-onderbrekingstaak meer beschikbaar zijn. Wanneer je visuele systeem vrij is en er een visuele afleiding is, dan word je afgeleid. Als er een kort moment in de taak is waarin je moet wachten, en er is iets anders dat je moet doen, dan word je afgeleid.

Externe onderbrekingen versus zelf-onderbrekingen

Wat klinkt beter, zelf kiezen wanneer je jezelf onderbreekt of worden onderbroken door een externe bron? Onze resultaten in Hoofdstuk 4 waren verrassend. Participanten waren langzamer in het afronden van de primaire taak (een geheugentaak met wiskundige vergelijkingen) wanneer ze zichzelf konden onderbreken dan wanneer ze extern werden onderbroken. Analyse van de pupilgrootte liet zien dat de pupil sterk vergrootte vlak voor een zelf-onderbreking door de keuze van de participant om zichzelf te onderbreken. Externe onderbrekingen waren daardoor vaak minder verstorend dan zelf-onderbrekingen. Echter, omdat mensen zichzelf over het algemeen onderbreken wanneer de werkdruk laag is, moeten de externe onderbrekingen ook plaatsvinden op momenten waarop de werkdruk laag is om ervoor te zorgen dat externe onderbrekingen minder verstorend werken dan zelf-onderbrekingen. In Hoofdstuk 5 van dit proefschrift hebben we dit resultaat gerepliceerd met een andere taak (emails beantwoorden). Participanten waren weer langzamer wanner ze zichzelf moesten onderbreken dan wanneer ze extern werden onderbroken. Ook hier reageerde de pupilgrootte anders op de twee verschillende types onderbrekingen.

In het tweede deel van mijn proefschrift heb ik gevonden dat een goed getimede externe onderbreking minder verstorend werkt dan zelf-onderbrekingen. Daarnaast heb ik pupilverwijding gebruikt om de effecten van externe onderbrekingen en zelf-onderbrekingen te vergelijken en ontdekt dat de keuze om jezelf te onderbreken een de pupil vergroot.

Onderbrekingsmanagementssysteem

Kunnen we de resultaten van dit proefschrift gebruiken om de werkomgeving efficiënter te maken? Omdat externe onderbrekingen tijdens moment met lage werkdruk minder verstorend zijn dan zelf-onderbrekingen, hebben we in Hoofdstuk 6 van dit proefschrift een onderbrekingsmanagementssysteem gemaakt dat mensen extern onderbrak op momenten van lage werkdruk. Het systeem nam de beslissing op welke moment er een lage werkdruk was door veranderingen in pupilgrootte, een bekende indicator van cognitieve werkdruk.

In het derde deel van dit proefschrift heb ik mijn eerdere resultaten gebruikt om minimaal verstorende externe onderbrekingen te maken. Het doel was om dit systeem te gebruiken om de tijdrovende beslissing om jezelf te onderbreken te voorkomen. Het lukte het systeem om externe onderbrekingen te maken op de juiste momenten, en de prestaties marginaal te verbeteren vergeleken met onderbrekingen op willekeurige momenten.

Conclusie

Onderbrekingen zijn een alledaags probleem die onze productiviteit beïnvloed. In dit proefschrift heb ik me vooral gericht op zelf-onderbrekingen die moeilijk te voorkomen zijn. Ten eerste heb ik gevonden dat mensen zichzelf onderbreken wanneer cognitieve middelen vrij zijn voor een onderbrekende taak. Ten tweede heb ik ontdekt dat goed getimede externe onderbrekingen minder verstorend zijn dan zelf-onderbrekingen vanwege de tijd die het kost om te beslissen om jezelf te onderbreken. Tenslotte heb ik een onderbrekingsmanagementssysteem ontwikkeld dat veranderingen in pupilgrootte gebruikt om de optimale momenten te vinden om mensen te onderbreken.

Acknowledgments

Finishing a PhD is hard. Having to move to a new country in order to do so, makes things even harder. There are a lot of people I have to thank for all the different kinds of help and support I received these four years.

First of all, I have to thank the "non-academic" people that made my life here easier and helped me focus on my work:

My morak, for moving to Nethetlands with me and making these four years better. Moving into a new country is hard, but sharing the responsibilities, the problems and the anxiety made this whole experience more easy and fun.

My parents for always being there, always helping me out and generally making my life so much easier.

Moving to the "academic" acknowledgements, I have to thank the following people:

First of all, Niels, for giving me the opportunity to do this PhD, for being one of the best supervisors and one of the brightest people in the world. I surely learned a lot from you.

Marieke, for patiently reading the same paper for the millionth time, for always being optimistic and for comforting me when I was getting rejected. It was always your "post-rejection" email that made me feel better.

My reading committee: Gloria Mark, Chris Wickens and Linda Steg, for taking the time to read my thesis and make useful comments.

My students, Tamara, Douwe, Maikel, Annemarie, Marlies, Patrick and Martijn, whose work helped my research a lot.

The cogmod group, Niels, Hedderik, Marieke, Jelmer, Fokie, Burcu, Harmen, Enkhbolt, Jacolien, Ben, Menno, Trudy, Chris, Marijke, Margreet, Florian, Udo, Tadeusz and everyone else that helped me over the years with useful feedback about my research and my presentation skills. However, the time of the meeting has to change, it is just cruel. Extra thanks to Menno for keeping me company during our trips and for helping me out a lot when I started working here. And to Trudy, for all the book talk, the sewing help and the office gossip.

Harmen of course also needs extra thanks for so many things: for keeping me company at zumba, for always helping out when needed and for being one of the nicest people I have met.

I also want to thank the AI PhD students for all the fun dinners we had. It was nice meeting all of you and eating all the food you cooked.

There are two people I have to thank even more. Without them, I don't know how I would have dealt with all the rejections, all the nights spent over my laptop and all the seemingly hopeless situations I had to face.

Jelmer, I am sure you know this thesis would not exist without you. My research and my mood would have been much worse if you were not part of this project. You taught me everything: how to make a cool presentation, how to talk to students, how to talk to professors, what this analysis means exactly, why my poster and dress should be pink, why it is important that my graph is exactly how you like it even if it takes me 2 more hours to make it. Just kidding, I don't know why the graph *should* be like that, but I know I have to do it. Apart from all the practical help, you always gave me hope, making the academic life seem easy and promising. I know I was annoying you all the time and I hope you found a way to enjoy our collaboration/friendship too. There is no way to thank you for everything you did for me, so you can just enjoy my eternal gratitude.

There is one more person that I cannot thank enough. I think of myself as a person that has a lot of good friends but doesn't really need a "best friend". And I certainly did not expect to find a best friend at my age (although I am still so much younger than you), a friend that we would even have our own "language" (daksi man, we are the best fo su!). Burcu, you know that I would be depressed if you were not here. You were here when we were having fun, you were here when I was complaining about everything and you were here when I was crying for stupid reasons. You are the most thoughtful person I know, you are literally always there when I need you (even when you have your own problems) and you always make me feel better (even if I don't always show it). I know I can be a very difficult friend, and thank you for being patient enough and not killing me. But I am sure you know how much you mean to me. You are the only person in my life that I know we will keep talking everyday even if we live in the opposite sides of the world. I don't know what your plan is, but you are stuck with me for the rest of your life. Accept it!

References

Adamczyk, P.D., & Bailey, B.P. (2004). If not now, when? The effects of interruptions at different moments within task execution. In *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2004*, 271-278

Adler. R.F., & Benbunan-Fich, R. (2013). Self-interruptions in discretionary multitasking. *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: Monsell, S., Driver, J.S.(Eds.), *Attention and Performance XVIII: Control of Cognitive Processes*. MIT Press, Cambridge, MA., 35-70

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

Altmann, E.M. & Trafton, J.G. (2007). Timecourse of Recovery from Task Interruption: Data and a Model *Psychonomic Bulletin and Review, 14*(6), 1079-1084.

Anderson, J.R. (1990). *The Adaptive Character of Thought.* Hillsdale, NJ: Erlbaum.

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 (1)*, 633-643.

Arroyo, E., & Selker, T. (2011). Attention and Intention Goals Can Mediate Disruption in Human-Computer Interaction. *Interact ‘11.*(6947), 454–470.

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,* 142–162, Cambridge, MA: CambridgeUniversityPress.

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.

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.

Bogunovich, P. & Salvucci, D.D. (2011). The effects of constraints on user behavior for deferrable interruptions. *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2011*

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*(2), 373-379.

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(2), 363-382

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 SIGHI Conference on Human Factors in Computing Systems: CHI 2015,* 2971-2980

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, 234–238.

Chiew, K.S., Braver, T.S. (2013). Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry. *Frontiers in Cognition.* 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(11), 1239–1243.

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

Czerwinski, M., Cutrell, E. & Horvitz, E. (2000). Instant Messaging and Interruption: Influence of Task Type on Performance, In Paris, C., Ozkan, N., Howard, S. and Lu, S. (Ed's.), *OZCHI 2000 Conference Proceedings*, 356-361.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2004*. 175-182

Dabbish, L., Mark, G., & Gonzalez, V.M. (2011). Why do I keep interrupting myself?: Environment habit, and self-interruption. In *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2011*, 3127-3130

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

Edward, M.B., & Gronlund, S.D. (1998). Task Interruption and its Effects on Memory. Memory, 6(6), 665–687.

Eyrolle, H., & Cellier, J.M. (2000). The effects of interruptions in work activity: field and laboratory results. *Applied Ergonomics*, 31(5), 537–543.

Fu, W.-T., & Gray, W.D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology, 52*(3), 195-242

Gillie, T. & Broadbent, D.E. (1989). What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, 50, 243-250.

Gonzalez, V.M., & Mark, G. (2004). “Constant, constant, multi-tasking craziness”: Managing multiple working spheres. *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2004*, 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*, 149 -153.

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: *Proceedings of the AAAI Conference on Human Computation & Crowdsourcing Works-in-Progress*. The Association for the Advancement of Artificial Intelligence (AAAI)

Granholm, E., Asarnow, R.F., Sarkin, A.J. & Dykes, K.L. (1996), Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33(4), 57–461.

Gray, W.D., Sims, C.R., Fu, W.-T., & Schoelles, M.J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*(3), 461-482.

Hess, E.H., & Polt, J.M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science,* 132, 349-350

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*, 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(1),* 103-115.

Hoeks, B. & Levelt W. (1993) Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods*, 25(1), 16–26.

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 SIGHI Conference on Human Factors in Computing Systems: CHI 2005*. 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 SIGHI Conference on Human Factors in Computing Systems: CHI 2005*. 1489-1492

Iqbal, S.T., & Bailey, B.P. (2006). Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2006*, 741–750.

Iqbal, S. T., & Bailey, B. P. (2007). Understanding and developing models for detecting and differentiating breakpoints during interactive tasks. In Proceedings of the ACM Conference on Human Factors in Computing Systems, 697–706.

Iqbal, S. T., & Bailey, B. P. (2008). Effects of intelligent notification management on users and their tasks. In *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2008*, 93–102.

Iqbal, S.T., & Bailey, B.P. (2010). Oasis: A framework for linking notification delivery to the perceptual structure of goal-directed tasks. *ACM ToCHI*, 17(4), 15:1-15:28.

Iqbal, S.T., & Horvitz, E. (2007). Disruption and recovery of computing tasks: Field study, analysis and directions. In *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2007*, 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 SIGHI Conference on Human Factors in Computing Systems: CHI 2004,* 1477-1480.

Jennings, J.R., & van der Molen, M.W. (2005). Preparation for speeded action as a psychophysiological concept. *Psychological Bulletin,* 131(3), 434–459.

Jin J. & Dabbish L.A. (2009) Self-Interruption on the Computer: A Typology of Discretionary Task Interleaving, In *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2009,* 1799-1808.

Kahneman, D. & Beatty, J. (1966). Pupil diameter and load of memory. *Science*, 154 (3756), 1583-1585

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. & 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.

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.

Kirchner, W.K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352-358

Kobayashi, Y., Tanaka, T., Aoki, K., & Fujita, K. (2015). Automatic Delivery Timing Control of Incoming Email based on User Interruptibility. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1779-1784.

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, 12(1), 13–21.

Laeng, B., Sirois, S. & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious?. *Perspectives on Psychological Science*. *7*(1), 18-27

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 451-468.

Lavie, N. (2010). Attention, Distraction and Cognitive Control under Load. *Current Directions in Psychological Science*, 19(3), 143-148.

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.M., & Harris, J. (2005). No task left behind? Examining the nature of fragmented work. *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2005*, 321-330

McFarlane, D.C. (1999). Coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction, INTERACT'99, IOS Press, Inc., The Netherlands*, 295-303.

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.

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

Nijboer, M., Taatgen, N.A., Brands, A., Borst, J.P. & van Rijn, H. (2013). Decision making in concurrent multitasking: Do people adapt to task interference? *PLoS ONE 8*(11): e79583. doi:10.1371/journal.pone.0079583.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 28, 411-421

Panepinto M.P. (2010). Voluntary Versus Forced Task Switching. In *Proceedings of the 54th Annual Meeting of the Human Factors and Ergonomics Society*, 453-457.

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

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

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 (2), 236-243.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

Rafiqi, S., Wangwiwattana, C., Fernandez, E., Nair, S., & Larson, E.C. (2015). Work-in progress, PupilWare-M: Cognitive Load Estimation Using Unmodified Smartphone Cameras. *MASS 2015,* SocialSens 2015.

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, 22(2), 204-207

Rogers, R.D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General,* 124(2), 207–231.

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. *Proceedings of the SIGHI Conference on Human Factors in Computing Systems: CHI 2010*, 85-88

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*, Santa Monica, CA: Human Factors and Ergonomics Society, 799-803

Salvucci, D.D., & Taatgen, N.A (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review*, 115, 101-130

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), 1819-1828

Salvucci, D.D., & Taatgen, N.A. (2011). The multitasking mind. *New York: 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, 1017-31

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, 17, 322-344

Sellen, A.J., Kurtenbach, G.P., and Buxton, W.A.S. (1990). The role of visual and kinesthetic feedback in the prevention of mode errors. *Proceedings of the Third IFIP Conference on Human-Computer Interaction, (INTERACT),* Cambridge, England, 667-673.

Spira, J.B., & Feintuch, J.B. (2005). The cost of not paying attention:

how interruptions impact knowledge worker productivity. *New York: BASEX.*

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.

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.

Tanaka, T., Abe, R., Aoki, K., & Fujita, K. (2015) Interruptibility Estimation Based on Head Motion and PC Operation. International Journal of Human-Computer Interaction, 31(3), 167-179

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 (5), 583–603.

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(12), e51134.

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 QuasiExperiment. *Journal of Applied Psychology, 97*(5), 1068-1076

Wickens, C.D. (2002). Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science, 3(2), 159-177.

Wierda, S.M., van Rijn, H., Taatgen, N.A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *PNAS*, 109(22), 8456–8460.

Züger, M., & Fritz, T. (2015). Interruptibility of Software Developers and its Prediction Using Psycho-Physiological Sensors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '15).

Publication List

Journal Articles

Katidioti, I., Borst, J.P., Van Vugt, M.K., & Taatgen, N.A. (in revision). Interrupt me: External Interruptions are less disruptive than self-interruptions.

Katidioti, I., Borst, J.P., Bierens de Haan, D. J., Pepping, T., Van Vugt, M.K., & Taatgen, N.A. (in revision). Interrupted by your Pupil: An Interruption Management System based on Pupil Dilation.

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

Peer-Reviewed Conference Papers

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.

Katidioti, I. Simpson, I. C., & Protopapas, A. (2012). Dual-route Connectionist model of Greek Spelling. In Proceedings ICCM 2012, Berlin, Germany

Abstracts

Katidioti, I., Borst, J.P., van Vugt, M.K. & Taatgen, N.A. (2015). Comparing self-interruptions with external interruptions and creating an interruption management system. In NVP Wintercongres 2015, the Netherlands

Katidioti, I., Borst, J.P., & Taatgen, N.A. (2014). Interrupt me! Why forced interruptions are better than voluntary interruptions. Psychonomic

Society's 55th Annual Meeting, Long Beach, CA, USA.

Katidioti, I., Borst, J.P., & Taatgen, N.A. (2013). What happens when we switch tasks: Pupil dilation in multitasking. In NVP Wintercongres 2013, the Netherlands

Katidioti I., Taatgen, N.A (2013). Be a better multitasker. How a pause in the primary task can turn a rational into an irrational multitasker. In Proceedings of the 35th Annual Meeting of the Cognitive Science Society, Berlin, Germany