University of Groningen

The Wandering Mind

Huijser, Stefan

DOI:
[10.33612/diss.198707305](https://doi.org/10.33612/diss.198707305)

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:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Huijser, S. (2022). *The Wandering Mind: investigating what drives self-generated thinking and determines its adaptive or maladaptive consequences*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.198707305>

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.

The Wandering Mind

Investigating what drives self-generated thinking and determines its adaptive or maladaptive consequences

Stefan Huijser

COLOPHON

This thesis was typeset with Adobe Indesign using the Candara font.

FUNDING

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

PRINTING

GildePrint Drukkerijen, Enschede, the Netherlands.

© S. Huijser, Deventer, The Netherlands, 2022.

university of
 groningen

The Wandering Mind

Investigating what drives self-generated thinking and determines
 its adaptive or maladaptive consequences

Proefschrift

ter verkrijging van de graad doctor aan de
 Rijksuniversiteit Groningen
 op gezag van de
 rector magnificus prof. dr. C. Wijmenga
 en volgens het besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op
 maandag 24 januari 2022 om 14:30 uur

door

Stefan Huijser
 Geboren op 27 januari 1993
 te Harlingen

Promotor

Prof. dr. N. A. Taatgen

Co-promotor

Dr. M. K. van Vugt

Beoordelingscommissie

Prof. dr. R. de Jong

Prof. dr. H. A. Slagter

Prof dr. J. Smallwood

CONTENTS

Chapter 1 Introduction	9
Chapter 2 The wandering self	19
Abstract	20
Introduction	21
Methods	28
Results	33
Discussion	42
Appendix	48
Chapter 3 Traces of ‘sticky’ thought	51
Abstract	52
Introduction	53
Methods	58
Results	68
Discussion	72
Chapter 4 Mechanisms of adaptive thought	81
Chapter 5 The art of planning ahead	93
Abstract	94
Introduction	95
Experiment 1	100
Methods	100
Results	109
Discussion	117
Experiment 2	119
Methods	120
Results	123
Discussion	131
General discussion	132
Chapter 6 Summary & conclusions	139
Nederlandse samenvatting	151
Acknowledgments	163

References **165**

List of publications **185**

An abstract painting with a textured surface, featuring a mix of blue, green, and brown tones. The brushstrokes are visible, creating a layered and organic feel.

Chapter 1

Introduction

In this chapter we give a short overview of this thesis and discuss the underlying theories and methodologies.

When I ask myself how I plan my day or how I would summarize my day, I notice that I focus a lot on tasks and events. For example, at the moment of writing (it is currently morning) I would say that I have brushed my teeth, ate breakfast, started writing on this introduction, and that I am planning to do the laundry in about an hour from now. While this ‘task-centric’ approach to plan or summarize a day is helpful and efficient, it provides a rather impoverished view on how the day actually looks like. Our waking lives are not merely experienced as sequences of tasks, but are also characterized by engaging in spontaneous, *self-generated*, thoughts that have little to do with our ongoing activities (Kane et al., 2017; Killingsworth & Gilbert, 2010). For example, I might interrupt myself to think about yesterday’s events while reading an article. Or, while writing this chapter, I may imagine what I will be doing in the upcoming weekend.

While self-generated thoughts occur frequently and are fundamental to our conscious experience, this topic has only been studied actively since the past 15 years. Similar to my task-centric approach to planning or summarizing my day, psychologists have traditionally held a very task-centric view on studying cognition. For example, an experiment designed to study a hypothesized working memory process will likely involve a task targeted at manipulating the hypothesized process. Any thought that occurs outside the scope of the task is either disregarded or else viewed as a distraction that needs to be controlled for.

The task-centric approach in psychology has been successful in advancing our knowledge on cognition. However, it only provides a limited perspective on how people actually engage in tasks, how they achieve their goals, and why they perform the way they do. To get a more complete picture, one needs to also understand the role of self-generated thought. Engaging in self-generated thought, of course, reduces attention towards what we’re currently doing resulting in decreased performance on that task. However, self-generated thought can also be beneficial to a task. It has been shown to be adaptive in aiding people to achieve personal goals by enhancing problem solving, creativity, and facilitating future planning (Baird et al., 2012; Baird, Smallwood, & Schooler, 2011; Mooneyham & Schooler, 2013).

With this thesis, I will try to get more insight in what role self-generated thought plays in how we engage in tasks, how we achieve our goals, and why we trade-off between our main task and task-unrelated thinking the way we do. I aim to provide a better understanding of the mechanisms that drive self-generated thought and the consequences of self-generated thought for ongoing and future task performance. However, before we delve into the work presented in this thesis, I first present some background on self-generated thought. I will outline a theoretical framework drives the

hypotheses tested and ideas posed in this thesis. This will be followed by an explanation of how I approached studying self-generated thought.

1.1 BACKGROUND AND METHODS

1.1.1 Theoretical framework

At first thought, you may think that the occurrence of self-generated thought is random due to its – oftentimes – spontaneous and unintentional nature. While there certainly is randomness in whether we engage in self-generated thought at a particular moment or not, many studies are indicating that there are systematic factors that influence its occurrence (McVay & Kane, 2013; Steindorf & Rummel, 2017). Research has shown that self-generated thought is more likely to occur in periods of low external demand (Forster & Lavie, 2009; Levinson, Smallwood, & Davidson, 2012; Rummel & Boywitt, 2014; Smallwood & Andrews-Hanna, 2013). Examples of such periods are task breaks – periods in which task processing temporarily ceases. Every experiment has task breaks, be it between stimuli, trials or blocks. Also outside of the lab task breaks are common occurrences, for example between reading the paragraphs of this introduction and in short pauses during conversations.

There are two reasons for why self-generated thought is more likely to occur during task breaks (Taatgen et al., 2021). First of all, task breaks imply that the ongoing activity is (likely) not occupying limited cognitive resources such as vision, working memory, motor functions etc. According to multiple resource theories of multitasking, two tasks can occur in parallel when they do not occupy or require the same resources at the same time (Salvucci & Taatgen, 2008; Wickens, 2002). Hence, self-generated thought can arise and continue when cognitive resources needed for its occurrence are available. For self-generated thought, these necessary cognitive resources are likely declarative memory for the retrieval of memories and working memory for storing intermediate representations (Fox, Nathan Spreng, Ellamil, Andrews-Hanna, & Christoff, 2015; Levinson et al., 2012; Smallwood & Schooler, 2006; Teasdale, Proctor, Lloyd, & Baddeley, 1993).

While the availability of required cognitive resources provides a suitable condition for self-generated thought to occur, this does not mean that it *will* occur necessarily. Processes from competing activities may take the opportunity instead. For example, in Chapter 2 I will present a study that involves a complex working memory (span) task. In between the presentation of the to-be-memorized items, participants will have the opportunity to engage in self-generated thinking. However, they may also decide to rehearse the items, which is arguably a more strategic choice. Therefore, a characteristic of task breaks is that they put you to the test. You are – metaphorically speaking – faced with the ‘decision’ to maintain your attention to the ongoing task or not. In this thesis, I

framed this ‘decision’ as a conflict resolution or competition process between so-called ‘mental operators’ of the ongoing task and self-generated thought. Mental operators are small rule-based processes that work towards a shared goal. For example, self-generated thought may be controlled by one mental operator that issues a retrieval from declarative memory, another operator that collects the retrieved memory and stores it in working memory, and a final operator that issues another retrieval with the previously retrieved memory as cue.

The principles underlying this competition between mental operators are derived from the Primitive Elements of Skill (PRIMs) cognitive architecture, which is both a theory of cognition as well as an architecture to simulate task processes (Taatgen, 2013). According to PRIMs, the probability of selecting and executing operators is dependent on whether the required resources are available and its activation relative to other operators. The activation of an operator is analogous to its current estimated relevance and usefulness. The activation of operators is determined, in part, by history of past use—when it has been used more frequently it is clearly more useful and therefore its activation is higher. On top of that, the activation of operators is also determined by context. Operators may receive activation from associated active goals in memory. For example, an active goal to read this chapter spreads activation to operators that support reading. Also, processing of cues in the external environment may spread activation to relevant operators. For example, seeing a phone spreads activation to operators for phone use.

The influence of context on the occurrence of self-generated thought has been suggested by several studies. For example, Steindorf and Rummel (2017) found that participants were more likely to use self-generated thought as an opportunity to rehearse items when they were told that they had to recall the items after the experiment compared to participants that were told that they did not need to recall the items later. Moreover, research on prospective memory (i.e., memory for postponed intentions) has demonstrated that spontaneous goal-oriented thinking is often triggered by internal cues (‘what happens to be in one’s thought’) and cues from our environment (see Kvavilashvili & Fisher, 2007; Szarras & Niedźwieńska, 2011).

1.1.2 Methodology

According to Smallwood and Schooler (2015), there are important core challenges that complicate investigating self-generated thought. First of all, self-generated thought is hard to experimentally induce due its, oftentimes, unintentional nature. Moreover, when you do have a working manipulation, the occurrence of self-generated thought is hard to capture as it produces little observable overt behavior. These challenges currently limit our capabilities in studying this phenomenon in depth and therefore also our

understanding of the thought process itself. Therefore, novel methodology is needed to investigate the process.

In order to study self-generated thought, two things are needed. The first is an experimental paradigm that provides opportunities for self-generated thought to occur. That is, the experimental paradigm should contain moments – task breaks – where self-generated thought can latch on available declarative and working memory resources to produce thoughts. Many studies on self-generated thought have used very simple and repetitive tasks in order to maximize the probability for self-generated thought to occur. However, as we will see in this thesis, *simple* is not an absolute requirement. Self-generated thought can also arise in more demanding contexts, as long as there are task breaks in which self-generated thought can compete with task processes for available resources.

Secondly, when self-generated thought occurs, one needs measurement techniques to capture it. Currently, the most frequently used method for capturing self-generated thought in experimental tasks is self-report. Participants are probed at various points during the task with questions asking whether their thoughts were related or unrelated to the task (Levinson et al., 2012; Weinstein, 2018). These short self-report questionnaires are commonly referred to as *thought probes*, and have shown to be reasonably valid in assessing the occurrence of self-generated thought (see McVay & Kane, 2012; however see Seli et al., 2015). Moreover, thought probes have the advantage that supplementary information on the mental content of self-generated thought can be easily gathered by including more questions. For instance, a study of Baird et al. (2011) included a question addressing the temporal orientation of the conscious experience (past or future) and found that self-generated thought contains mostly future-oriented thoughts. In Chapter 2 and 3 of this thesis, I will also use thought probes to capture self-generated thought during the experiment. I will include additional questions to assess the occurrence of particular types of self-generated thought.

While thought probes have been used effectively in many studies, they do have disadvantages. Thought probes rely on subjective judgements, lack temporal precision, and provide only limited insight in the mechanism of self-generated thought and its consequences to ongoing and future performance. Also, the introduction of thought probes may alter the nature of self-generated thought itself. Therefore, I will combine self-report with measures of task performance and eye tracking in this thesis. Combining, or *triangulating between*, these measures allows for identifying patterns in task performance and eye tracking measures that correspond to differential thought probe reports (Franklin, Smallwood, & Schooler, 2011; Smallwood, Brown, et al., 2011; Unsworth

& Robison, 2016). In turn, these identified patterns may inform our understanding of the mental processes that underly self-generated thought.

Measuring task performance allows for investigating interference between self-generated thought and the ongoing activity. Examining *how* and to *what extent* self-generated thought and an ongoing task interfere with each other can tell something about what happens when we engage in self-generated thought. For example, Seli, Cheyne and Smilek (2013) combined thought probe reports and measures of task performance to investigate whether self-generated thought makes cognitive control for the purpose of the immediate task impossible (see e.g., Smallwood & Schooler, 2006). Participants in the study performed a metronome response task, in which they were asked to make timed responses at the pace of repeating rhythmic tones. Good performance in this task requires controlled processing to ensure consistent responding. The results of the study showed that in the five trials preceding reports of self-generated thought, response times were more variable compared to the same period prior to on-task reports. Therefore, this study provided evidence for the idea that self-generated thinking draws controlled processing resources away from an ongoing task (for similar findings see e.g., Bastian & Sackur, 2013; Jubera-García, Gevers, & Opstal, 2019).

Another way to examine the effect of self-generated thought is through eye tracking. Eye tracking is a technique that allows for continuous recording of where people are looking on a screen (i.e., *eye movement*) and how large their pupil is at current moment (i.e., *pupil size*). Eye movement is a useful measure because the location of a person's gaze is a reasonable indicator of where the person's attention is directed towards and consequently, what his or her goals are. Taking this further, sequences of gazes at locations that are relevant to a particular task or activity are a likely indicator that a person is engaged in the task. For example, when you see someone gazing at different products on a supermarket shelf, it is reasonable to infer that this person is looking for a particular product. When expected eye movement behavior, given a task, is not observed, this may indicate that a person is engaged in self-generated thought. For this reason, patterns in eye movement will be used as an (indirect) indicator of self-generated thought in this thesis.

Another measure that has been associated with self-generated thought is pupil size. Pupil size is typically measured on two temporal scales. The first of these is the task-evoked response in pupil size. The task-evoked response is a transient increase in pupil size following the processing of a stimulus (Hoeks & Levelt, 1993). The magnitude of this response is currently understood to reflect the amount of attention, cognitive control, and cognitive processing required by the task (Gilzenrat, Nieuwenhuis, Jepma, & Cohen,

2010; Unsworth & Robison, 2014; Wierda, Van Rijn, Taatgen, & Martens, 2012). Hence, measuring task-evoked responses in pupil size allows for investigating how variance in task probe reports corresponds to variance in the amount of attention or cognitive control exerted for the immediate task.

Pupil size is also measured during task breaks, at which point it is referred to as baseline pupil size. Baseline pupil size is proposed to reflect overall arousal levels and the tendency to seek for novelty (Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Murphy, Robertson, Balsters, & O'Connell, 2011; Unsworth & Robison, 2016). Large baseline pupil size has been correlated with a state of over-arousal and a tendency to explore new behaviors. Smaller baseline pupil sizes have been related to under-arousal and inactivity. In this thesis, I will measure both task-evoked responses in pupil size and baseline pupil size. In this thesis, task-evoked responses are used to investigate how different mental states influence the amount attention exerted to the ongoing task, while baseline pupil size is used to investigate how different arousal states correspond to different mental states, as reported with thought probes.

1.2 OVERVIEW OF THIS DISSERTATION

As mentioned in the first section of this Introduction chapter, the aim of this thesis is to provide more insight into the workings of self-generated thought. In addition, I would like to provide more understanding of when, how, and to what extent self-generated thought has maladaptive consequences for the ongoing task and when it has adaptive consequences to future tasks.

Central to this thesis are the following questions:

- How does competition for goals and cognitive resources affect the likelihood of engaging in self-generated thought?
- How can self-generated thought be traced continuously by means of behavioural and eye tracking measures?
- How and why is self-generated thought adaptive?

This thesis will be organized in two parts. In the first part, consisting of Chapter 2 and 3, I will explore the maladaptive side of self-generated thought and examine its mechanisms. Chapter 2 presents an experiment in which I investigated whether processing information related to oneself can trigger distracting self-generated thought. This experiment tests the predictions of a cognitive model, which describes that self-generated thought can arise from competition between task and distracting processes. In Chapter 3, I present an eye tracking experiment examining how more difficult-to-disengage-from self-generated thought impairs task performance and is visible with pupillary measures

(both baseline and task-evoked). The second part explores the adaptive side of self-generated thought. In Chapter 4, I present a literature review that reviews how and why self-generated thought can support adaptive behavior. Finally, I will examine in Chapter 5 to what extent we can engage in adaptive self-generated thought while being under various degrees of load from our immediate task.

Chapter 2

The wandering self

In this chapter we investigated how self-referential processing affects self-generated thought to test the predictions of a computational cognitive model.

This chapter has been previously published as:

Huijser, S., van Vugt, M. K., & Taatgen, N. A. (2018). The wandering self: Tracking distracting self-generated thought in a cognitively demanding context. *Consciousness and Cognition*, 58, 170-185.

ABSTRACT

We investigated how self-referential processing (SRP) affected self-generated thought in a complex working memory task (CWM) to test the predictions of a computational cognitive model. This model described self-generated thought as resulting from competition between task- and distracting processes, and predicted that self-generated thought interferes with rehearsal, reducing memory performance. SRP was hypothesized to influence this goal competition process by encouraging distracting self-generated thinking. We used a spatial CWM task to examine if SRP instigated such thoughts, and employed eye tracking to examine rehearsal interference in eye movement and self-generated thinking in pupil size. The results showed that SRP was associated with lower performance and higher rates of self-generated thought. Self-generated thought was associated with less rehearsal and we observed a smaller pupil size for mind wandering. We conclude that SRP can instigate self-generated thought and that goal competition provides a likely explanation for how self-generated thoughts arises in a demanding task.

Keywords: self-generated thought, mind wandering, self-referential processing, task demand, computational cognitive modeling, eye tracking

2.1 INTRODUCTION

When we try to characterize our conscious experience, we can say that is guided in roughly two ways (see Dixon, Fox, & Christoff, 2014). On one hand, our thoughts are shaped by what we perceive in our external environment. For instance, when we are reading a book, our thoughts are likely on what we read. On the other hand, our conscious experience is not always tied to our environment. Analogously, we might find ourselves thinking about future plans while reading this article. Or: we might contemplate about the answer to a question we just received. In these situations, constructive and associative processes in memory shape our thoughts internally without the need of perceptual guidance.

This internally directed thought, or *self-generated thought*, has been extensively studied in past recent years with examples including: thinking about the self or others, mental simulations, imagination, past and future thinking, and planning (for a review see Andrews-Hanna, Smallwood, & Spreng, 2014). Self-generated thought can occur both in the absence and the presence of external stimuli. In addition, it can be triggered both by internal and external influences, but it always involves a disengagement from the external environment (Schooler et al., 2011). In this article, we will focus on self-generated thought that occurs in the context of a main external task but is not relevant to performance on this task. In other words, we want to explore self-generated thought that becomes a distraction from the main task.

One popular example of self-generated thought, oftentimes regarded as a distraction from main activities, is *mind wandering*. Mind wandering refers to a self-generated thought process that proceeds independently of an external task (Smallwood & Schooler, 2015). Since mind wandering is covert in nature, it can only be observed indirectly with self-report questions that sample thought content (see Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004). These self-report questions, or *thought probes*, are embedded in the task to determine how mind wandering is related to performance on the task. Studies using thought probes have identified that mind wandering is associated with higher and more variable response times (e.g., Bastian & Sackur, 2013; van Vugt & Broers, 2016) and to decrements in task performance (e.g., working memory; Rummel & Boywitt, 2014). Although mind wandering has clear adverse consequences for our current activities, this does not need to be true for tasks still in the future. The content of mind wandering has been found adaptive in many instances, with thoughts that are highly self-focused, goal-directed, and future oriented (Baird et al., 2011).

Many studies have examined mind wandering in tasks imposing low cognitive demand, allowing cognitive resources for mind wandering to occur. However, one might argue that it is more interesting to investigate mind wandering (and also other self-generated

thoughts) in contexts where one is busily engaged in a task, situations where performance on main activities is at stake. Research has shown that the adverse consequences of mind wandering are greater in tasks with higher cognitive demand (such as in n-back and general intelligence measures; see McVay & Kane, 2012a, 2012b; Mrazek et al., 2012; Rummel & Boywitt, 2014; Smallwood & Andrews-Hanna, 2013). Commonly, it is observed that mind wandering is less frequent in higher demand contexts (see e.g., Smallwood et al., 2011; Thomson, Besner, & Smilek, 2013), but proportions are still substantial. Research by Seli and colleagues (2016) even found equal rates of mind wandering across task demand, with relatively more reports of unintentional mind wandering in a challenging task and more intentional mind wandering in an easier task. Similar to what has been found for lower demand contexts, the content of mind wandering is predominantly self-related and focused on personal goals and future plans (Smallwood, Schooler, et al., 2011). Consequently, these results suggest that mind wandering, and potentially self-generated thought in general, has similar characteristics in contexts of low and high cognitive demand.

In the present study, we will focus on demanding task paradigms to examine self-generated thought in a context where it becomes a distraction for a main task. Building on earlier work (viz., Daamen, Vugt, & Taatgen, 2016), we are interested in investigating the potential of self-referential processing to instigate such thoughts. Our approach is to test the assumptions of a computational cognitive model, which describes the cognitive processes of participants in a task and predicts how self-generated thought can occur in short spare moments of a task. We will test whether self-generated thought occurs in spare moments of the task by means of triangulating eye tracking measures and thought probes. However, before explaining the full details of our experiment, we will first provide a short review on correlates of self-generated thought in eye tracking measurements. Subsequently, we will discuss the role of self-referential processing in instigating self-generated thought.

2.1.1 Correlates of self-generated thought in eye tracking

The thought probes we just introduced have the advantage to provide insight in the content of self-generated thought. However, the discrete nature of the measure makes it hard to determine when the thoughts occurred. Correlates in physiological or task performance measures (e.g., response times) can provide this knowledge and may allow us to track self-generated thought online (e.g., Franklin, Smallwood, & Schooler, 2011; Grandchamp, Braboszcz, & Delorme, 2014; Mittner et al., 2014). Here we will review studies using eye movement and pupil size measurement, since these are the measures we will use in our current study.

With eye tracking we continuously measure eye movement and pupil size throughout an experiment while having the advantage of being entirely unobtrusive (unlike e.g., fMRI and EEG). Eye movement correlates of self-generated thought have commonly been investigated in linguistics studies, in which this measure is employed to examine real-time cognitive processing during reading (e.g., Reichle, Reineberg, & Schooler, 2010; Uzzaman & Joordens, 2011). For instance, a recent study by Foulsham, Farley, and Kingstone (2013) found that participants showed longer and slightly more frequent fixations in sentence reading during mind wandering. According to the researchers, the longer fixation duration are a result of a decoupling between seeing the words and processing their meaning, which provides evidence for the idea that self-generated thought, such as mind wandering, involves a disengagement of attention from the external task (see Kam & Handy, 2013).

Similar results were found in a recent study by Benedek and colleagues (2017). The researchers examined the influence of self-generated thought (called internally-directed cognition in their study) on eye movement and pupillometric measures by manipulating the focus of attention (internal and external) with an externally demanding task (sentence generation) and an internally demanding task (anagram). The rationale was that the anagram task would require self-generated thought to solve, building on an internal representation of the problem. The sentence generation task, however, could be solved without. The researchers found that self-generated thought was associated with fewer fixations and saccades, but with increased fixation duration. Based on the results, the researchers conclude that eye movements are more spontaneous and less directed during self-generated thought. In addition to eye movement differences, they also identified a larger average pupil size for self-generated thought, indicating higher levels of arousal (see Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Murphy, Robertson, Balsters, & O'Connell, 2011). Potentially, the higher levels of arousal reflected the decoupling of attention from the external task, which is hypothesized to occur with self-generated thought.

Research on pupil size correlates for mind wandering has not been fully consistent. Similar to the above-mentioned study of Benedek, some studies have identified larger pupil sizes prior to reports of mind wandering (Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Smallwood et al., 2012; Smallwood, Brown, et al., 2011), whereas other studies have found smaller pupil sizes (e.g., Grandchamp, Braboszcz, & Delorme, 2014; Konishi, Brown, Battaglini, & Smallwood, 2017; Mittner et al., 2014; Unsworth & Robison, 2016). Explanations for these mixed results are speculative. Notable about the above mentioned studies is that larger pupil sizes were primarily found in relatively undemanding tasks (e.g., reading) whereas the smaller pupil sizes were identified in

more demanding tasks (e.g., n-back). In more difficult tasks, mind wandering may result in a lower arousal state compared to on-task thoughts, resulting in a relatively smaller pupil size. Easier tasks are known to engender more future oriented and intentional mind wandering, which may be associated with higher levels of arousal.

In conclusion, self-generated thought is not unequivocally associated with patterns in eye movement and pupil size. Tasks without external stimuli show less but more spontaneous eye movements, whereas tasks with higher visual demand (e.g., reading) show more frequent eye movements. Self-generated thought seems to be associated with smaller pupil sizes in task with relatively high cognitive load, but with larger pupil sizes in tasks with lower load.

2.1.2 Prior work: Instigating self-generated thought in a CWM task

In the present study, we want to extend on earlier work (viz., Daamen et al., 2016), in which we assessed self-generated thought in a demanding task and demonstrated that when participants were asked to think about themselves, performance on a memory task decreased compared to when they made simple semantic judgments. We came to this understanding by using a variation on a verbal complex working memory task (CWM). In such a task, participants have to memorize items (i.e., letters) while also performing a processing task between each presented item. We assumed that the processing task blocked any rehearsal processes. Therefore in order to do well in this task, a good strategy would be to squeeze in rehearsals whenever this was possible. One moment during which rehearsal would be possible was a 1-second blank period that followed the processing task. However, if the processing task evoked distracting self-generated thoughts, these may have been pursued in the blank instead.

The frequency of self-generated thought was manipulated by introducing a self-referential processing (SRP) task in the paradigm. Processing self-related information is known to evoke highly automatic elaboration processes in autobiographical memory (see e.g., Conway & Pleydell-Pearce, 2000; Conway, Singer, & Tagini, 2004; Wessel et al., 2014), which might also trigger mind wandering through spontaneous associations with e.g. personal goals or aspirations (see Barnhofer, Crane, Spinhoven, & Williams, 2007; D'Argembeau et al., 2007; Klinger & Cox, 2004). We contrasted the self-referential processing condition with a control condition in which participants had to decide whether a depicted object fitted in a shoebox. The results of our study showed worse memory performance for the self-referential processing condition compared to the shoebox condition.

Figure 2.1. Overview of the Primitive Elements of Skill (PRIMs) model of self-generated thought in a complex working memory task. WM refers to working memory. Decl. Memory refers to declarative memory/long-term memory. The self-generated thought processes are mediated by the 'wander' and 'elaborate' operators (in dark blue).

2.1.3 Computational cognitive model of distracting self-generated thought

We predicted that the reduction in memory performance in the self-referential processing condition was the result of increased elaboration of the processing words and mind wandering during the blank periods. To formalize this prediction, we previously built a computational cognitive model of the task in the PRIMs cognitive architecture (Taatgen, 2013; see also Hiatt & Trafton, 2015; van Vugt, Taatgen, Sackur, & Bastian, 2015, for other models of mind wandering). PRIMs describes cognition in terms of interactions and exchange of information between multiple cognitive resources (i.e., vision, motor, working memory, long-term memory, and control). This information sharing process is controlled via operators, which are small rule-based operations that work towards a shared goal. For instance, our model (see Figure 2.1) has an operator that stores a target item in working memory, and another operator to retrieve it for recall. A fully working

model requires several of these operators to complete a task and uses a selection process to find the right operator at the right time. This selection process is based on the priority (i.e., an ‘activation’ value) of the operator and can be influenced by the content of the cognitive resources. Competition between operators during the selection process and the influence of the content of the cognitive resource in this competition, are the key mechanism of our model.

Following existing literature on the influence of processing self-related information and memory systems (see Self-Memory System theory; Conway & Pleydell-Pearce, 2000), we expected that the task could elicit two ‘types’ of self-generated thought. First of all, self-generated thought may appear as *mental-elaboration* on the word stimuli from the processing task, which means that participants will resume thinking about their decisions. These mental elaborations are created by spontaneous elaboration processes in memory, which have been implicated with self-referential processing (e.g., Conway & Pleydell-Pearce, 2000; Conway et al., 2004; Wessel et al., 2014). And secondly, the self-related material may instigate *mind wandering* by means of evoking spontaneous associations with e.g. concerns or personal goals. Both types of self-generated thought are triggered by the task, but rely on internal representations in memory for their continuity. The content of mental-elaboration is mostly related to the task, whereas mind wandering is task-unrelated. Irrespective of task-(un)relatedness, we argue that both are irrelevant to performance on the task since engaging in them during blank periods will interfere with rehearsal of the targets.

Despite the fact that we expected two different manifestations of self-generated thought, our current model does not distinguish between different types. The reason for this is that we were primarily interested in modeling when and why self-generated thought may occur, and not so much about modeling the phenomenology of the thought processes.

We modeled both manifestations of self-generated thought with the same two operators that control the elaboration on the words from the processing task and possible further associations. These operators do not belong to the task itself, and have therefore little priority. However, when self-related information enters the cognitive resource of vision and/or the working memory, it can temporarily increase the priority of operators that control self-generated thought. Eventually, the model will start to engage in self-generated thought because the priority of its operators will have become high enough to allow them to be selected. The chances of this happening are lower for the control condition because the words during this processing task do not increase the priority of the operators for self-generated thought since they do not prompt the elaboration

process (dark blue shapes in Figure 2.1).

So how can this model explain the performance decrement for the self-referential processing condition? As described above, the verbal CWM task has a short 1-s break (i.e., blank period) following the processing task. During this break, we assumed that participants would rehearse the items of the CWM task, as rehearsal was difficult to perform in the rest of the trial. However, it could also be an opportunity for e.g. mind wandering. The model behaves similarly. During the break, the model can choose between its task goal to rehearse the items or to select self-generated thought operators. Because these operators have a higher priority in the self-referential processing condition, the model is more likely to engage in self-generated thought during this spare moment in the task. Consequently, the model does not rehearse and will most likely forget some target items. The result is that the model will perform relatively worse when it is experiencing self-generated thought, thereby producing the performance decrement observed in the behavioral data.

2.1.4 Current study

The main goal of the present study is to investigate whether self-referential processing can influence the propensity to self-generated thoughts that are irrelevant to performance in a demanding complex working memory task. For this purpose we tested the assumption of our computational cognitive model, which posits that self-referential processing is a strong competitive process, capable of triggering self-generated thought during spare moments of the task (i.e., blank periods) at a cost for rehearsal. We included thought probes to measure thought directly, allowing us to examine whether participants were rehearsing the targets (i.e., *on-task*) or not (i.e., *off-task*) and whether self-generated thinking increased in the self-referential processing condition. We also used eye tracking to provide additional measures to the thought probes. By combining both self-report and physiological measures we could give a more objective account of self-generated thought in the blank periods of the task, without being solely dependent on one measuring technique.

We re-designed the verbal CWM from Daamen et al. (2016) into a spatial version with locations as target items instead of letters, allowing us to track rehearsal (or the absence thereof) with an eye-tracker. Memory researchers have found that rehearsal of spatial locations is accompanied with eye movements to these locations (see e.g., Logie, 1995). We assumed that rehearsals mediated by eye movements would take more time to perform compared to the verbal rehearsals in Daamen et al. (2016). Therefore, we also extended the duration of the blank period from 1 second to 2 seconds. Fixations to locations of the to-be-memorized targets were interpreted as an effort to rehearse the

targets. We hypothesized that there would be fewer fixations on target locations for the self-referential processing condition compared to control, reflecting that self-generated thought (i.e., mental-elaboration and mind wandering) interfered with rehearsal (see Benedek et al., 2017).

With respect to the pupil size measurements, we hypothesized that episodes of mind wandering would be associated with smaller pupil sizes (see Grandchamp, Braboszcz, & Delorme, 2014; Konishi & Smallwood, 2016; Mittner et al., 2014; Unsworth & Robison, 2016). We had no clear hypothesis about the pupil size for incidences of mental-elaboration. Since we expected more mind wandering to occur in self-referential processing trials compared to control trials, we hypothesized we would find smaller average pupil sizes in self-referential processing trials compared to control.

For both eye movement and pupil size, we focused on the blank periods following the processing task. These periods were expected to show most rehearsal and self-generated thought as it was predicted by the model to show most competition since the task does not give either of the two goals a clear preference.

2.2 MATERIAL AND METHODS

2.2.1 Participants

A total of 38 native Dutch speakers participated in this study. All participants were screened for having normal or corrected-to-normal vision and received a small monetary compensation as reward. We excluded the data of one participant from the analysis as this participant performed close to chance level on the processing task in the control condition, which suggests that this participant did not perform the processing task seriously. Another five participants were removed after we found through visual inspection of the eye tracking data that there was excessive data loss (i.e., > 20 % of missing data and loss of data in multiple consecutive trials). This left us with a final sample of 32 participants (18 female, $M_{\text{age}} = 22.4$, $SD_{\text{age}} = 2.6$). All experiments were conducted in accordance with the Declaration of Helsinki, and all participants signed an informed consent form prior to the start of the experiment.

2.2.2 Apparatus

The participants were tested in a small windowless room at the University of Groningen. This room contained a chair, a desk on which the computer and the eye-tracker were located, and a chin-rest attached to the desk. The experiment was presented on a 20 inch LCD-monitor (1600 x 1200, 64 PPI) and was positioned approximately 60 cm from the chin-rest (SR-Research Head Support).

Figure 2.2. Overview of one trial in the spatial version of the complex working memory task used in this experiment. The trial in this figure represents a self-referential processing condition trial. The diagonal arrow shows the progression of a trial in time.

In order to measure eye movement and pupil size, we used an Eyelink 1000 eye-tracker from SR Research, which was capable of measuring gaze position at a spatial resolution of $< .001^\circ$ RMS and pupil size at 0.2% of the diameter. The sampling rate was set at 250 Hz.

The stimulus presentation was controlled with PsychoPy (Peirce, 2007). Before the start of the experiment, a 9-point calibration was performed using the software of the eye-tracker. We performed no additional calibrations during the experiment, unless the eye-tracker frequently lost track of the pupil. Re-calibration for this reason was necessary for 5 of our participants, of which two participants were in the end excluded from analysis (the two participants are included in the five mentioned in section 2.2.1 *Participants*).

2.2.3 Materials and design

In the interest of tracking mind wandering and rehearsal, we examined pupil size and eye movement (fixations) in a spatial variant of the complex working memory task (see Figure 2.2). Our task required participants to memorize the locations of target items (an 'X', Helvetica font, 40 pixels in height) in a 4x4 matrix in sequential order, while also performing a processing task after each presented target item. Similar to our previous work (Daamen et al., 2016), the processing task involved evaluating words corresponding two separate conditions. In the control condition, nicknamed 'shoebox', participants were presented with object words (Helvetica font, 30 pixels in height) in the same 4x4 matrix and decided whether the depicted object fitted in a shoebox. All the selected words in this condition were written in Dutch and contained no examples of objects with negative valence (e.g., knife, needle, bullet). The experimental self-referential condition

required the participants to evaluate whether a presented personality trait (e.g., kind) described them ('Am I kind?'). For both conditions we used the same word set as the study of Daamen.

We used a grey background (120,120,120; rgb255) for the entire experiment including calibration. The stimuli and instructions were presented in black (0,0,0; rgb255) and the matrix surrounding the stimuli had an intermediate color (85,85,85; rgb255) between the grey background and the black stimuli.

2.2.3.1 Thought probe question

Throughout the experiment, the participants were periodically presented with thought probes following recall (see Figure 2.2). These thought probes were used to sample thought content during the blank period prior to recall.

The question we used was a variant of the thought probe introduced by Stawarczyk, Majerus, Maj, Van der Linden, and D'Argembeau (2011) and Unsworth & Robison (2016). The question was (translated from Dutch), 'What were you thinking about before you were prompted to answer?', with the following response options: (1) I tried to remember the location of the X's; (2) I was still thinking about the words from the decision task (= processing task); (3) I was evaluating aspects of the task (e.g., my performance, how long it takes, difficulty of the task); (4) I was distracted by my environment (sound/temperature etc.) or by my physical state (hungry/thirsty); (5) I was daydreaming/ I thought about task-unrelated things, (6) I was not paying attention, but I did not think about anything specific.

Consistent with other research (e.g., Stawarczyk et al., 2011), we labeled response option 3 as *task-related interference*, option 4 as *external distraction*, option 5 as *mind wandering*, and option 6 as *inattentiveness*. We labeled option 2 as *mental-elaboration*. Response option 1 was included to sample if participants were rehearsing during the blank period and was labeled as *on-task* thought in this study. We refer to all the other response options (i.e., excluding on-task) combined as *off-task*.

2.2.4 Procedure

When participants entered the eye tracking room, they received instructions on the procedure of the experiment and were asked to sit down in front of the computer and eye-tracker. Following the instructions, the chin-rest was adjusted to the height of the participant and a calibration of the eye-tracker was performed. After calibration, the experiment started.

The participants were first presented with a reminder on the decision rules in the processing task (decision rules: ‘Does this object fit in a shoebox’ or ‘Does this word describe me?’) and the corresponding response mapping (press ‘m’ for yes and ‘z’ for no). After instructions, one block of 4 practice trials followed. Each trial started with an instruction on the decision rule required in the processing task and a reminder to the response mapping. This instruction remained on the screen until the participants pressed the space bar. Thereafter, a fixation cross was shown at the center of the screen for a random interval between 800 and 1200 milliseconds. Following the fixation cross, the first storage item was presented in the matrix at a random and unique (not previously occupied within a single trial) position for 1 second. This allowed the participants to encode the stimulus. Subsequently, a 4 seconds self-paced processing task started with the presentation of a processing word at a random location in the same matrix. The locations of the words were random and changed every second from position to position to ensure interference with rehearsal during the processing task itself. After processing, the matrix was cleared and remained empty on the screen for 2 seconds. During this blank phase, participants could rehearse the items (or engage in any other cognitive activity, since there was no instruction). The storage phase, processing phase, and blank were repeated a number of times equal to the span (3 or 4). Thereafter, the trial ended with a prompt to recall the item locations in sequential order. On practice trials, feedback was presented after recall showing performance on recall and the processing task.

In this within-subject design, all participants performed both conditions with half of the trials being ‘shoebox’ trials (48 trials) and the other half being self-referential processing trials. The 96 trials were distributed across 6 blocks (16 per block), with an equal amount of trials for both conditions and span in each. Thought probes were equally but randomly distributed across the experiment (48 in total, 8 per block). The order of the trials within the blocks was randomized for every participant. The practice block was similar in its structure to the experimental blocks except for the fact that it contained a total of 4 trials and these trials were only of span 3. Completing the experiment took approximately an hour.

2.2.5 Preprocessing of the eye tracking data

To prepare the eye tracking data for analysis, we performed several preprocessing steps. First, we removed all the automatically detected blinks from the data including 100 milliseconds before and after the blink events. In addition, sudden downward jumps in pupil size (> 200 units in pupil size in one sample, comparable to 0.5 SD in pupil size) were regarded as artifacts and subsequently removed. The discarded data were linearly interpolated and the resulting full data set was down-sampled to 100 Hz.

Following down-sampling, we segmented the data into trials with only storage, processing, and blank periods (removing the inter-trial intervals). To calculate the average pupil size in the blank periods, we selected the pupil size measurements recorded during fixations, thereby leaving out saccades (see also Unsworth & Robison, 2016). Given the spatial nature of our task and the hypothesized differences in eye movement behavior, we subsequently corrected the pupil size measurements for gaze direction with linear regression (for details about this method see Brisson et al., 2013). The linear regression model quantified the relationship between the X- and Y-coordinates of the eye gaze and the pupil size. Because the eye-tracker system recorded pupil size in arbitrary units and individual participants had different mean pupil sizes, we thereafter z-score normalized the pupil size data within participants (see also Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Smallwood, Brown, et al., 2011; Unsworth & Robison, 2016). The average pupil size for the different conditions and probe responses was then computed by averaging the z-score normalized pupil size for each individual blank period.

As a proxy for rehearsal activity, we determined the number of correct fixations in each blank period. We corrected this for how many fixations a participant made by dividing it by the total amount of fixations counted in the blank period. Correct fixations were defined as fixations on locations where previously in a trial a storage item was presented. We checked whether any drift in the measurement of X- and Y-coordinates affected the integrity of the rehearsal activity score. Drift in the measured coordinates of the eye gaze could have led to incorrect classification of the fixations' location (with locations we refer here to the cells in the 4x4 matrix). Correcting of the coordinates for drift in the experiment with linear regression did not change the classified location of the fixations. Therefore, drift in the measurements did not affect the rehearsal activity score.

2.2.6 Statistical analysis

All analyses were performed on trials followed by a thought probe (i.e., half of the data). This made sure that tests on condition and thought probe responses were done on identical data sets. We refer to the *Appendix Table 1* for all the test results on the full data set. Results from the full data set did not change most of the conclusions in this paper.¹

The main dependent variable in the behavioral analysis was memory performance on the task. We expressed memory performance as a percent correct score, computed by dividing the number of correct responses by the span for each trial. Average pupil size and rehearsal activity (see previous section) were used as dependent variables in the eye

¹ The results were the same as on half of the data set, except for the effect of condition on pupil size. Condition was significant in the full data set, but not in the reduced data set. However, since the effect size was small in both data sets, we conclude that the true effect of condition on pupil size was only marginal.

tracking analysis. Predictors of interest included condition (self-referential processing, control), span (3, 4), block (1 to 6), thought probe response (1 to 6), attentional state (on-task, off-task), and response times in the processing task (in seconds).

Hypotheses in the behavioral analysis and eye tracking analysis were tested with chi-squared likelihood ratio tests comparing an intercept model with a model containing only the factor of interest. Contrast testing was performed with Tukey Post-Hoc tests. Since memory performance and rehearsal activity were expressed as proportions, we fitted them as generalized linear mixed effect models with Laplace approximation, assuming a binomial distribution with a logit link function. Models with pupil size or response times as dependent variable were fitted as maximum likelihood linear mixed effects models.

We used the R statistical programming language (www.r-project.org, version 3.2.4) to perform all the statistics. For fitting the linear mixed effect models and generalized models we used the 'lmer' and 'glmer' function from the lme4 package (version 1.1-12).

2.3 RESULTS

2.3.1 Behavioral results

The main goal of the behavioral data analysis was to investigate if self-referential processing (SRP) decreased memory performance on the task. With that, we also wanted to verify whether participants reported more often to be off-task (i.e., not rehearsing the targets) in self-referential processing trials compared to control trials, and whether this difference existed because of an increase in self-generated thoughts (i.e., mental-elaboration and mind wandering).

The results showed that participants performed well on this task with a mean percent correct score of 0.82 (SD = 0.28) or 82% correct on a trial. This score is similar to that observed in other studies adopting similar tasks (see e.g., Redick et al., 2012). Performance did not change with blocks ($\chi^2(5) = 8.028$, $p = 0.15$), implying that there was neither a learning effect nor a fatigue effect in our experiment. An alternative explanation could be that the two canceled each other out. Although performance was unaffected by blocks, participants did show significantly lower performance for span 4 trials ($M_{\text{diff}} = -0.04$, $SD_{\text{diff}} = 0.09$; $\beta = -0.306$, $SE = 0.078$; $\chi^2(1) = 15.714$, $p < 0.001$), demonstrating that a higher span was indeed more difficult.

In agreement with our previous study, we found a significant performance decline for self-referential processing compared to the control condition ($M_{\text{diff}} = -0.05$, $SD_{\text{diff}} = 0.09$; $\beta = -0.398$, $SE = 0.076$; $\chi^2(1) = 27.401$, $p < 0.001$; Figure 2.3A). There was no interaction effect between condition and span ($\beta = -0.136$, $SE = 0.156$; $\chi^2(1) = 0.765$, $p = 0.38$), suggesting

Figure 2.3. Mean percent correct separated by span and condition (A) and separated by attentional state and condition (B). SRP = self-referential processing, control = shoebox. On-task = reports of rehearsal of the targets, off-task = reports of thoughts not related to rehearsal. Error bars represent one standard error of the mean.

that the effect of condition on memory performance was similar for both spans. We also did not find an interaction between condition and attentional state (i.e., on-task or off-task thinking) on memory performance ($\beta = 0.232$, $SE = 0.169$; $\chi^2(1) = 1.870$, $p = 0.17$; Figure 2.3B). This was expected, because rehearsal or the absence of it should

have similar effects on memory performance in both conditions. When we tested for main effects of condition and attentional state (i.e., in one model) in the absence of this interaction, we found that being off-task significantly attenuated performance on top of what could be explained by condition ($\beta = -1.021$, $SE = 0.090$; $\chi^2(1) = 135.44$, $p < 0.001$). As Figure 2.3B clearly shows, the impact of off-task thinking on memory performance ($\beta = -1.134$, $SE = 0.123$) was much greater compared to that of condition. Nonetheless, condition also predicted memory performance ($\chi^2(1) = 7.679$, $p = 0.005$), but the size of the effect was reduced with the added effect of attentional state (from $\beta = -0.398$, $SE = 0.076$; to $\beta = -0.219$, $SE = 0.079$). This shows that condition and attentional state both explain unique variance in memory performance on the task, but also that a part of the effect of condition is explained by differences in on- and off-task thinking.

Necessary to note is that in our previous study, we found that response times to the processing words (i.e., time to make each decision) were higher in self-referential processing trials and that increasing response time negatively affected memory performance. Similar results were found in the present study. With increasing response times we found a decrease in performance on the task ($\beta = -0.930$, $SE = 0.164$; $\chi^2(1) = 31.57$, $p < 0.001$). The average response time for the semantic processing decisions (e.g.,

‘Does *marble* fit in a shoebox?’) in the control condition was 99.2 ms (SD = 23.5 ms), whereas the self-referential processing decisions (e.g., ‘Am I *kind*?’) took on average 115.9 ms (SD = 30.7 ms). This difference in response time was found to be significant ($\beta = -0.167$, SE = 0.010; $\chi^2(1) = 249.42$, $p < 0.001$). These results indicate that the increase in response time in the self-referential processing condition was not used to rehearse the targets, otherwise an improvement in memory performance would be expected.

A possible explanation for the longer response times with self-referential processing, is that the processing was more cognitively demanding compared to the semantic processing in the control condition. Therefore, an alternative explanation for the observed performance decrement is that higher cognitive demand (i.e., task difficulty) when performing self-referential processing, reflected in longer response times, caused the lower memory performance. We reanalyzed the effect of condition on memory performance with an added fixed effect for the response times in the processing task. This allowed us to test whether there was an effect of condition while assuming that response times were constant. We found that self-referential processing still predicted lower memory performance when we controlled for response times ($\beta = -0.268$, SE = 0.083; $\chi^2(1) = 10.265$, $p = 0.001$), but the effect did become smaller (from $\beta = -0.398$ to $\beta = -0.268$). When we also added attentional state to the model, condition was not significant anymore ($\beta = -0.112$, SE = 0.086; $\chi^2(1) = 1.668$, $p = 0.19$), but we did find an effect for attentional state ($\beta = -1.007$, SE = 0.090; $\chi^2(1) = 130.41$, $p < 0.001$) and a relatively smaller effect for response times ($\beta = -0.571$, SE = 0.184; $\chi^2(1) = 9.402$, $p = 0.002$). Taken together, these results show that the observed performance decrement can most likely be explained by differences in how often the participants were not rehearsing the targets (i.e., off-task), and in part by the higher difficulty of the self-referential processing task.

2.5.1.1 Thought probe results

Analyzing the distribution of thought probe responses revealed that participants were off-task in 51.0% of the trials. This means that participants reported having other thoughts than rehearsal of the targets on half of the analyzed trials. The proportion of off-task reports was higher in the self-referential processing condition ($\chi^2(5) = 87.379$, $p < 0.001$; see Figure 2.4A), with 60.2% off-task in self-referential processing trials and 41.8% in control trials. Most importantly, post-hoc testing showed that participants experienced more self-generated thought with more *mind wandering* ($\beta = 0.938$, SE = 0.214; $z = 4.381$, $p_{\text{adjusted}} < 0.001$) and more *mental-elaboration* ($\beta = 1.164$, SE = 0.138; $z = 8.441$, $p_{\text{adjusted}} < 0.001$) when processing self-referential material. We found that in the self-referential processing condition the number of mental-elaboration reports doubled (from $M = 14.1\%$ to 30.7%), and that mind wandering nearly doubled (from $M = 5.1\%$ to 8.9%) in comparison with the control condition. Having off-task thoughts during a trial

Figure 2.4. Overview of the mean frequency of different thought probe categories in the self-referential processing (SRP) and control condition (A) and memory performance as mean percent correct for each thought probe category (B). Error bars represent one standard error of the mean.

was found to negatively impact task performance ($\chi^2(5) = 122.71$, $p < 0.001$; see Figure 2.4B) and similar post-hoc testing showed that all different thought probe reports were associated with lower performance on the task (all $p_{\text{adjusted}} < 0.001$).

Similar to the previous section, where we analyzed the effect of condition under simultaneous statistical control for attentional state and response times, we found that the different thought probe reports ($\chi^2(5) = 172.53$, $p < 0.001$) and response times ($\beta =$

Figure 2.5. Overview of rehearsal activity during blank periods for on-task vs. off-task trials (A; all thought probe responses combined except for on-task) and the entire distribution of thought probe responses (B). Error bars represent one standard error of the mean.

-0.582, $SE = 0.185$; $\chi^2(1) = 9.647$, $p = 0.002$) reliably affected memory performance but not condition ($\beta = -0.102$, $SE = 0.088$; $\chi^2(1) = 1.668$, $p = 0.24$). Interestingly, post-hoc testing revealed that both *mental-elaboration* ($\beta = -0.840$, $SE = 0.109$; $z = -7.734$, $p_{\text{adjusted}} < 0.001$) and *mind wandering* ($\beta = -1.447$, $SE = 0.153$; $z = -9.485$, $p_{\text{adjusted}} < 0.001$) and the other reports (all $p_{\text{adjusted}} < 0.001$) significantly decreased performance. Taken together, the results show that self-referential processing increased the frequency of self-generated thought (i.e., mental-elaboration and mind wandering) and that the occurrence of self-generated thought was detrimental to performance on the task.

2.3.2 Eye tracking results

The aim of the eye tracking data analysis was two-fold. First of all, we wanted to objectively determine whether there was less rehearsal activity in self-referential processing trials, in trials reported as off-task in general, and specifically in trials with self-generated thought. If so, this would provide support for our model mechanism claiming that rehearsal is more often replaced by another process (presumably self-generated thought) in the self-referential processing condition. Furthermore, we wanted to examine the influence of condition and different thought probe reports on the average pupil size, which is thought to be smaller for self-referential processing and – across both conditions – for trials with mind wandering.

2.3.2.1 Eye movement - rehearsal

Our analysis showed an average rehearsal activity of 0.48, meaning that on average 48% of the fixations during blank periods were on locations where previously a target was presented. This is substantially more than would be expected by chance if eye

movements were unrelated to rehearsal and therefore random. In that case the rehearsal activity would be 0.12, which is the average probability of accidentally looking at a target location combined for span 3 and span 4 trials. The average rehearsal activity performed by the participants varied substantially ($SD = 0.37$), with averages as high as 0.67 and as low as 0.29. The correlation between rehearsal activity and memory performance on the task was small ($r = 0.36$, $df = 30$, $p = 0.045$) and also showed substantial variation (95% CI [0.01, 0.63]). The variability in average rehearsal activity and in its relationship with memory performance reflects that not all participants used and/or benefited as much from eye movements to perform rehearsal.

Consistent with our prediction that self-referential processing would interfere with rehearsal, we found a difference between conditions in rehearsal activity ($\beta = -0.076$, $SE = 0.027$; $\chi^2(1) = 8.174$, $p = 0.004$). This difference was however very small with an average difference of 2% in correct fixations ($SD_{diff} = 7\%$). Together with the weak correlation ($r = 0.36$) between memory performance and rehearsal activity, this suggests that variation in (overt) rehearsal may not explain why we see a decrement in memory performance between the conditions.

However, as we have seen in the previous paragraph, not all participants relied on overt rehearsal. Therefore, we also analyzed the effect of condition on rehearsal activity while taking individual differences in the relationship between rehearsal activity and memory performance into account. We computed a correlation coefficient between each individuals' rehearsal activity and their memory performance to obtain a measure of 'reliance' on eye movements to memorize the targets for each participant. We tested for an interaction between condition and reliance on eye movements for rehearsal, and found that rehearsal activity decreased in self-referential processing trials with increasing reliance ($\beta = -0.493$, $SE = 0.153$; $\chi^2(1) = 10.315$, $p = 0.001$). In other words, participants that used overt rehearsal and benefited from using it showed a stronger decrease in rehearsal in self-referential processing trials.

Alongside differences between conditions, we were also interested in the effects of self-generated thought and other off-task thought on rehearsal activity. We predicted that when participants reported to have other thoughts than rehearsing the targets (i.e., off-task), we would find a decrease in rehearsal activity measured in eye movement. Our results confirmed this prediction ($\beta = -0.154$, $SE = 0.029$; $\chi^2(1) = 28.349$, $p < 0.001$; see Figure 2.5A). The effect of attentional state was not found to significantly differ between conditions ($\beta = -0.042$, $SE = 0.056$; $\chi^2(1) = 0.566$, $p = 0.45$).

Analyzing the individual thought probe responses in more detail, we found that all the

Figure 2.6. Mean pupil size in a single span four trial. S = storage phase (i.e., presentation of new target item), P = processing phase, B = blank phase.

responses collectively impacted rehearsal activity ($\chi^2(5) = 37.742$, $p < 0.001$; see Figure 2.5B). Post hoc testing revealed that only *mind wandering* ($\beta = -0.221$, $SE = 0.058$; $z = -3.789$, $p_{\text{adjusted}} = 0.002$), *mental-elaboration* ($\beta = -0.185$, $SE = 0.036$; $z = -5.203$, $p_{\text{adjusted}} < 0.001$), and *inattentiveness* ($\beta = -0.215$, $SE = 0.009$; $z = -2.896$, $p_{\text{adjusted}} = 0.04$) were associated with less rehearsal activity. Therefore, not all off-task thinking impacted rehearsal activity

equally. As can be seen from Figure 2.5B, trials with task-related interfering thoughts had very similar rehearsal activity scores. Trials where participants experienced external distraction showed a lower but also more variable average rehearsal activity.

Since the response times in the processing task were found to differ between conditions and to affect memory performance, we also analyzed the influence of these response times on rehearsal activity. We expected to observe less rehearsal when more effort is put in processing of the word stimuli. Our results confirmed that indeed longer response times predicted less rehearsal activity ($\beta = -0.215$, $SE = 0.036$; $\chi^2(1) = 36.819$, $p < 0.001$). Under simultaneous statistical control of individual thought probe reports (results were similar for attentional state) and response times, we found no evidence anymore for an effect of condition on rehearsal activity ($\beta = -0.002$, $SE = 0.028$; $\chi^2(1) = 0.006$, $p = 0.94$). However, individual thought probe reports ($\chi^2(1) = 27.839$, $p < 0.001$) and response times ($\beta = -0.194$, $SE = 0.037$; $\chi^2(1) = 28.006$, $p < 0.001$) did significantly impact how much rehearsal activity was present in the blank periods. Post-hoc testing showed that only the self-generated thought states *mental-elaboration* ($\beta = -0.161$, $SE = 0.034$; $z = -4.367$, $p_{\text{adjusted}} < 0.001$) and *mind wandering* ($\beta = -0.205$, $SE = 0.056$; $z = -4.367$, $p_{\text{adjusted}} < 0.001$) reliably influenced rehearsal. Taken together, we conclude that the observed difference in rehearsal activity between the conditions is likely explained by an increase in self-generated thought and by increased difficulty of processing in the processing task.

Figure 2.7. Overview of the average pupil size during the blank phase by condition (A) and across time by condition (B). Error bars represent one standard error of the mean.

2.3.2.2 Effects of the task on pupil size

Prior to analyzing the average pupil size, we first wanted to examine the dynamics of the pupil size within a trial. Since little is known about pupil size over the course of a complex working memory task, we first examined pupil size over the whole task (Figure 2.6). As can be seen from the figure, the individual phases (i.e., storage ‘S’, processing ‘P’, and blank ‘B’) are easily distinguishable. The processing phases yielded the largest pupil size, followed by the storage phase and the blank phase. As the relative size of the pupil is known to be positively correlated with mental effort (see e.g., Beatty, 1982), this could mean that the processing task required most effort. This ordering in the phases is to be expected, as the storage phase only involved memorizing one extra target, and the blank phase did not require anything except for optional rehearsal.

In this study we focused our hypotheses on the blank phases, therefore we calculated the average pupil size in these parts of the trial. Unlike many other mind wandering studies where the average is calculated from a relatively steady pupil size, our period shows a substantial decrease in the size of the pupil from its onset (Figure 2.6). This decline attenuated as the trial progressed, suggesting that magnified storage load also relatively increased exerted mental effort. Despite the observed decline, we believe that this does not affect the validity of our average pupil size measures. As can be seen in Figure 2.7B, the difference in the average pupil size is very similar throughout the blank phase. Furthermore, the difference in average pupil size was found to not differ between phases across condition ($\chi^2(3) = 0.997$, $p = 0.80$), across different thought probe responses ($\chi^2(3) = 16.439$, $p = 0.36$) and attentional state ($\chi^2(3) = 2.513$, $p = 0.47$).

Figure 2.8. Overview of the average pupil size during the blank phase by on-task vs. off-task (A) and across separate thought probe categories (B). Error bars represent one standard error of the mean.

2.3.2.3 Average pupil size in blank periods

After analyzing the task dynamic effects of the task on the pupil size, we first examined the relationship between the average pupil size during blank periods as an indicator of alertness/arousal and condition (Figure 2.7A,B). We expected to find a smaller average pupil size for the self-referential processing condition due to increased mind wandering.

As can be seen in Figure 2.7, the average pupil size behaved in the hypothesized way with a relatively smaller pupil size during blank periods in self-referential processing trials. However, this difference was not significant ($\beta = -0.039$, $SE = 0.023$; $\chi^2(1) = 2.500$, $p = 0.11$). The manipulation did not significantly affect the average pupil size in blank periods, but how did the pupil react to off-task thinking? Our analysis showed that the average pupil size was lower in off-task trials compared to on-task trials ($\beta = -0.089$, $SE = 0.024$; $\chi^2(1) = 13.363$, $p < 0.001$), suggesting a lower state of alertness and arousal when participants reported not to rehearse the targets (Figure 2.8A). Similar to rehearsal activity, we found no evidence for an interaction effect between condition and attentional state on the average pupil size ($\beta = 0.051$, $SE = 0.050$; $\chi^2 = 1.059$, $p = 0.30$).

We also analyzed the distinct influences of different off-task reports on the average pupil size (see Figure 2.8B). Confirming our prediction in line with Unsworth's study, we found that *mind wandering* was associated with a smaller average pupil size ($\beta = -0.266$, $SE = 0.049$; $z = -5.406$, $p_{\text{adjusted}} < 0.001$). In addition, a similar effect was found for a state of *inattentiveness* with no specific recollection of thought ($\beta = -0.337$, $SE = 0.063$; $z = -5.338$, $p_{\text{adjusted}} < 0.001$). In contrast to Unsworth, the average pupil size for *external distraction* was smaller instead of larger ($\beta = -0.251$, $SE = 0.052$; $z = -4.830$, $p_{\text{adjusted}} < 0.001$). This difference can potentially be explained by the nature of our task. Our task

was much more visually demanding than the sustained attention task in Unsworth's study, and given that the pupil responds to both arousal and mental effort, one potential explanation is that a distraction from the task resulted in lower arousal and therefore a smaller average pupil size. From both we can expect a lower arousal state and hence a smaller pupil size. For both *mental-elaboration* ($\beta = 0.046$, $SE = 0.031$; $z = 1.503$, $p_{\text{adjusted}} = 0.64$) and *task-related interference* ($\beta = -0.077$, $SE = 0.039$; $z = -1.951$, $p_{\text{adjusted}} = 0.35$) we found no significant difference in pupil size with on-task.

Longer response times during the processing task were found to increase the average pupil size in the blank periods ($\beta = 0.233$, $SE = 0.029$; $\chi^2(1) = 58.985$, $p < 0.001$). Since larger pupil sizes are commonly associated with increased mental effort or resource allocation, this supports the idea that the self-referential processing was more cognitively demanding compared to the semantic processing in the control condition. Controlling for response times by adding it as a fixed effect in our linear mixed effects models did not change any of the preceding conclusions.

2.4 DISCUSSION

In agreement with earlier work (viz., Daamen et al., 2016), we showed that performance on a spatial complex working memory task was worse when participants had to perform self-referential processing compared to semantic decision-making (i.e., control condition). This consistency between studies with different stimulus sets indicates that the observed performance decline is robust and reliable. Our results suggested that there were two main factors that contributed to the observed performance decline: distracting processes (such as self-generated thought) that interfered with memorizing the targets, and increased cognitive demand for the self-referential processing task.

From the thought probe reports, we observed that participants reported less rehearsal of the targets in blank periods following self-referential processing (49.8% vs. 58.2% for control). Instead we found higher incidences of self-generated thought, with twice as many reports of mental-elaboration on the processing material (14.1% to 30.7%), as well as an increase in mind wandering (5.1% to 8.9%). The increase of self-generated thought provides confidence in our cognitive model, which assumes that self-related material can increase the likelihood of processes that issue elaborative thought, and that such processes can be initiated at spare moments in the task.

Increased reports of mental-elaboration with self-referential processing are in line with existing theories about self-reference in autobiographical memory (Barnhofer et al., 2007; Conway & Pleydell-Pearce, 2000; Conway, Singer, & Tagini, 2004; Wessel et al., 2014). According to Conway et al. (2004), the main function of autobiographical memory

is to provide an accurate window on our current personal goals and values such that we can behave in a goal driven manner. With that, we try to maintain a sense of self that is consistent with these goals and values (called ‘self-coherence’). When participants processed self-referential material in our experiment, the stimuli might have given rise to thoughts about events that conflicted with their sense of self. Conway and colleagues suggested that this self-discrepancy automatically triggers elaborative thought to resolve this conflict, distracting us from our initial memory search. The mental-elaboration that we observed in our study might therefore reflect an effort of participants to resolve discrepancies of the self with their current goals and values. Interestingly, mental-elaborations in our study occurred outside of the context of the processing task, therefore drawing attention away from the task to memorize the targets in favor of self-generated thinking. Our study therefore suggests that mental elaborations can be pervasive.

It is interesting that also mind wandering occurred more often in trials involving self-referential processing. We hypothesized that processing self-related material would increase the amount of mind wandering, because it might trigger associations in memory with e.g., concerns that are unrelated to the task. The increase in our study provides some support for this hypothesis. Still, reports of mind wandering were infrequent (5.1% for control, 8.9% for self-referential processing). Although the amount of mind wandering did almost double from the control condition, we do think that given the small amounts of observed mind wandering, more research is warranted before we can conclude that self-referential processing is a good manipulation for mind wandering.

Alongside thought probe reports, we also examined participants’ eye-fixations and pupil size to obtain a more fine-grained picture on how self-referential processing and self-generated thought influenced cognitive processes during the blank periods. Eye-fixations were used to determine overt rehearsal activity by computing what percentage of eye-fixations were on locations where previously in a trial a target was presented. Our results showed that such fixations occurred much more often than chance (48% instead of 12%) and that rehearsal activity measured in this way was positively correlated with memory performance ($r = 0.36$). At the same time, the limited magnitude of the correlation indicates that not all participants needed eye-fixations to memorize the targets. Consequently, we argue that our measure did reflect overt rehearsal, but that it could not capture all rehearsal activity. In particular rehearsal from covert memorization strategies (i.e., without eye movements) were likely not reflected in the measure (see also Godijn & Theeuwes, 2012).

Consistent with this proposal, we also observed less overt rehearsal in blank periods of

trials where participants experienced self-generated thought. This suggests that self-generated thought could interfere with rehearsal in the spare moments of the task, providing some support for our model assumption that self-generated thought replaces rehearsal processes. We furthermore expected to find a similar reduction in rehearsal activity in self-referential processing trials because of self-generated thought. Our results indeed indicated a reduction, but the size of this effect was very small ($M = -2\%$ in rehearsal). Considering that not all participants relied as much on overt rehearsal to memorize the targets we did find a stronger effect for participants that benefited more from overt rehearsal. For that reason, we argue that self-referential processing indeed could interfere with rehearsal in the blank periods of the trial. For participants that relied on overt rehearsal, this likely also attenuated performance.

Our pupil size results showed that different manifestations of self-generated thought were differentially associated with the size of the pupil. In agreement with many earlier studies using cognitively undemanding task (see e.g., Grandchamp, Braboszcz, & Delorme, 2014; Mittner et al., 2014; Unsworth & Robison, 2016), our results indicated that mind wandering was associated with a smaller pupil size. A smaller pupil size as correlate for mind wandering may therefore also generalize to demanding paradigms. The smaller pupil size during blank periods reflected that mind wandering occurred in a state of lower arousal and alertness, providing additional support for our model prediction that self-generated thinking disrupted task-focus during spare moments. Unlike mind wandering, mental-elaborations were not found to be associated with a different pupil size in comparison with efforts to rehearse the targets (i.e., on-task thought). Conway and Pleydell-Pearce (2000) argued that mental-elaboration involves strategic retrieval processes that require control processes. Therefore, this potentially suggests that mental-elaboration required similar amounts of control processes compared to efforts to memorize the targets, resulting in a similar average pupil size.

It should be noted that next to self-generated thought also cognitive demand of the processing task was an influential factor. Self-referential processing appeared to be more demanding, with higher average response times compared to control. Higher response times negatively impacted memory performance and rehearsal activity, and increased pupil size. Therefore, all the results are at least in part explained by the fact that the self-referential processing task was more difficult to perform. Nonetheless, all our analyses indicated that even after controlling for response time, an effect of self-generated thought remained, suggesting that cognitive demand in the self-referential processing condition is not the only factor causing differences in memory performance.

To summarize, the results and discussions above suggest that processing self-related

material could increase the amount of self-generated thoughts in the task at moments when they prevent memory rehearsal. The negative impact of such lingering thoughts on memory performance, attenuation in pupil size (only for mind wandering), and reduction in goal-driven eye movement behavior in spare moments of the task, likely indicates that self-generated thought interfered with the task goal to memorize the targets. This study therefore provided some support for our computational cognitive model assuming that distracting self-generated thought occurs as a result from competition between processes of the task (i.e., rehearsal) and processes that cause self-generated thoughts. In addition, we found support that self-referential processing can leverage the competition towards self-generated thought.

2.4.1 New insights on self-generated thought in demanding tasks

With increased confidence that self-referential processing could increase self-generated thought, and that self-generated thought could interfere with memorizing the targets, we think it is important to discuss what this study can tell us about self-generated thought in demanding contexts.

First of all, this study indicates that self-generated thought is not restricted to undemanding tasks, but that it is also pervasive in demanding contexts, even when engaging in it is irrelevant to performance on the task. Based on our cognitive model and existing research, we believe that the occurrence of such thought depends on two important factors. First of all there *needs to be an opportunity for self-generated thought* to start i.e., there should be no ongoing task processes that prevent off-task thinking from occurring. In other words, the necessary cognitive resources for self-generated thought need to be available. This is in line with resource theories of mind wandering, which claim that mind wandering is a process that requires resources to start and to continue (see e.g., Levinson, Smallwood, & Davidson, 2012; Smallwood & Schooler, 2006). In addition, it is consistent with existing multitasking theories, in particular that of threaded cognition (Salvucci & Taatgen, 2008, 2010). Threaded cognition assumes that cognition involves multiple cognitive resources (e.g., vision, control, motor, retrieval) and that various tasks in a multitasking setting require different or similar resources at a given time. When tasks ask for different resources, they can be carried out in parallel. This account predicts that people switch tasks when the resources for another task are available. That this happens was shown by a recent task-switching study of Katidioti and Taatgen (2014), who observed that participants switched between the main task and an interrupting secondary task when faced with a short delay in the main task (i.e., forced delay of 3 seconds). The delay periods in their study parallel with the blank periods in the present experiment. Therefore, this would predict that the blanks provided an opportune moment in our task to switch to thinking such as elaborative thoughts or

mind wandering. The results of this study indicated that self-generated thinking in general indeed existed in the blank periods, and given that blanks were frequent, this can explain why self-generated thought could occur frequently in our CWM task despite the fact that cognitive demands were high.

The second important component is that processes leading to self-generated thought have to compete with task processes when opportune moments occur. This competition involves selecting either task related processes or self-generated thought processes, and can be influenced by the relative priority of such processes at a given time. On the surface, this idea is consistent with the control failure theory of mind wandering (McVay & Kane, 2009), which explains that mind wandering occurs when we fail to maintain task focus. Our model, however, does not require any control processes to explain this, as the selection process is fully automatic. When resources for mind wandering are available, processes from different tasks and goals will compete. The relative priority of these processes will determine what will be thought about next. The results of our study suggested that self-referential processing may be such a strongly competitive process.

2.4.2 Implications for aptitude measurement

The results of this study could have implications for measuring working memory capacity in aptitude assessment. In research, but also in many applications outside of research, we measure working memory capacity and similar aptitudes to predict someone's success in various situations (Conway et al., 2005). Why working memory capacity has predictive utility in many different contexts is unknown, however the present study suggests that the ability to maintain task-focus is an important factor (see e.g., Kane, Hambrick, & Conway, 2005). We found that self-generated thought and other distractions significantly impacted working memory performance, therefore failures to maintain task focus were an important predictor of performance. Similar results were found in a recent study of Mrazek et al. (2012). The researchers determined that mind wandering behavior explained 49% of the variation in performance on a working memory test, general intelligence test, and a scholastic aptitude test.

From the above we could conclude that the ability to maintain focus on our current activities is likely an important component of aptitude tests. Although this may be an important ability, we believe that caution is needed. In our study, we were able to manipulate the prevalence of self-generated thought and therefore the measured working memory performance. This shows that context can have a substantial influence on measuring working memory capacity. When people have for example pressing concerns or experience conflicts with their sense of self, this could increase self-generated thought and that could in turn decrease the measured working memory

capacity (see Klinger & can affect the relative priority of the current task goal and other goals. Therefore, in order to explain internally-directed behavior in demanding tasks, one has to consider both how natural task breaks and the availability of resources can allow it to arise, and how likely one can maintain task focus in such situations.

2.5 CONCLUSIONS

In this study we aimed to give more insight in how and why we engage in self-generated thought in a cognitively demanding task, and how self-referential processing influences self-generated thinking. The results of this study and the model of the task suggested that distracting self-generative thoughts in demanding tasks can be pervasive and that they might arise when cognitive resources are available, such as when there are natural breaks in current ongoing activities. How likely and how often self-generated thought occurs potentially depended on the relative priority of the current task to other goals that compete for attention. Processing self-related information could bias the priority of self-generated thought, but one could also think that other processes such as boredom or fatigue can affect the relative priority of the current task goal and other goals. Therefore, in order to explain internally-directed behavior in demanding tasks, one has to consider both how natural task breaks and the availability of resources can allow it to arise, and how likely one can maintain task focus in such situations.

APPENDIX

	β	SE	χ^2	df	p
<i>Behavioral</i>					
Condition	-0.312	0.054	33.589	1	< 0.001
Span	-0.309	0.055	31.840	1	< 0.001
Block	0.004	0.016	0.063	1	0.80
Condition x Span	0.007	0.110	0.004	1	0.95
Response times (RT)	-1.101	0.119	85.22	1	< 0.001
Condition (with RT)	-0.132	0.059	5.032	1	0.02
<i>Average pupil size</i>					
Condition	-0.035	0.017	4.137	1	0.04
Response times (RT)	0.285	0.021	169.12	1	< 0.001
Condition (with RT)	-0.093	0.018	27.242	1	< 0.001
<i>Rehearsal activity</i>					
Condition	-0.049	0.019	6.713	1	0.010
Response times (RT)	-0.153	0.025	37.254	1	< 0.001
Condition (with RT)	-0.023	0.020	1.347	1	0.25

Table 1. Overview of test results for the full data set. In the first column, the fitted main effects and interaction effects are shown for the model with behavioral data (proportion correct per trial), average pupil size data, and rehearsal activity. Models were fitted as linear mixed effects models (LMEs) with XX as random effects.

Chapter 3

Traces of ‘sticky’ thought

In this chapter we explored the ‘stickiness’ dimension of thought, by investigating how stickiness is reflected in task performance and pupil size.

This chapter has been previously published as:

Huijser, S., Verkaik, M., van Vugt, M. K., & Taatgen, N. A. (2020). Captivated by thought: “Sticky” thinking leaves traces of perceptual decoupling in task-evoked pupil size.

PLoS ONE, 15(12), e0243532.

ABSTRACT

Throughout the day, we may sometimes catch ourselves in patterns of thought that we experience as rigid and difficult to disengage from. Such “sticky” thinking can be highly disruptive to ongoing tasks, and when it turns into rumination constitutes a vulnerability for mental disorders such as depression and anxiety. The main goal of the present study was to explore the stickiness dimension of thought, by investigating how stickiness is reflected in task performance and pupil size. To measure spontaneous thought processes, we asked participants to perform a sustained attention to response task (SART), in which we embedded the participant’s concerns to potentially increase the probability of observing sticky thinking. The results indicated that sticky thinking was most frequently experienced when participants were disengaged from the task. Such episodes of sticky thought could be discriminated from neutral and non-sticky thought by an increase in errors on infrequent no-go trials. Furthermore, we found that sticky thought was associated with smaller pupil responses during correct responding. These results demonstrate that participants can successfully report on the stickiness of their thought, and that stickiness can be investigated using pupillometry. In addition, the results suggest that sticky thought may limit attention and exertion of cognitive control to the task.

Keywords: sticky thought, self-generated thought, pupil size, task performance, thought probes, sustained attention, perceptual decoupling

3.1 INTRODUCTION

3.1.1 Background

In response to pressing concerns and unreachd goals we may catch ourselves in thoughts that we feel are difficult to disengage from. For example, we may be absorbed in thinking about a recently received paper rejection, while we should actually be reading this article. In general, task-unrelated thought is frequently referred to as *mind wandering* (Smallwood & Schooler, 2015). However, in cases such as the paper rejection, these thoughts may not leave us alone, and make it very difficult to concentrate on our immediate tasks. In this case, one can call these thoughts *sticky* (Joormann, Levens, & Gotlib, 2011; van Vugt & Broers, 2016). An extreme form of such sticky thought is rumination, a rigid and narrow-focused thought process that is hard to disengage from and often negative in valence and self-related (Nolen-Hoeksema & Morrow, 1991). In general, rumination causes individuals to be unable to concentrate and devote their attention to tasks at hand because attention is focused internally instead (van Vugt & Broers, 2016). However, in contrast to depressive rumination, sticky thoughts could also have a positive valence, for example when we are caught up in a pleasant fantasy that we do not want to let go of, or thoughts with desire for a delicious cookie keep recurring in our minds (Kavanagh, Andrade, & May, 2005; Papies, Stroebe, & Aarts, 2007). Another term for sticky thought is perseverative cognition. Perseverative cognition has been associated with activation of the physiological stress system, and has been proposed to play a key role in the onset and maintenance of depression (Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008) and anxiety (Barlow, 2002; Nolen-Hoeksema, 2000). Finally, sticky thought is closely related to the concept of *constrained thinking* (Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016; Mills, Raffaelli, Irving, Stan, & Christoff, 2018). Constrained thinking refers to an experience in which thoughts do not move freely but instead are focused on a narrow set of content. It is different from our concept of sticky thinking in the question that is posed to the participant—while sticky refers to the experience of the participant that it is difficult to drop the current stream of thought, constrained refers to participants' experience of having a stream of thought that is – deliberately or not – restricted to a narrow set of content.

Yet, sticky thoughts—especially in their non-clinical form—could also have advantages to the individual. By temporarily shielding thought from external distractions, they can help the individual to work on future goals (Kam et al., 2011; Marchetti, Koster, Klinger, & Alloy, 2016; Mooneyham & Schooler, 2013; Smallwood, Brown, et al., 2011). When goals remain unattained and concerns unresolved but the thoughts remain, sticky thoughts may become increasingly more intrusive, disrupting our everyday functioning (Joormann et al., 2011; Klinger, 2013; van Vugt & Broers, 2016). Therefore, sticky thoughts may have important effects on our performance in everyday tasks.

Sticky thought has mostly received attention in literature on psychopathology. Studies have demonstrated that perseverative cognition has measurable negative effects on somatic health (for a review see Verkuil, Brosschot, Gebhardt, & Thayer, 2010). For example, rumination and worry are associated with prolonged activation of the immune system (Denson, Spanovic, & Miller, 2009), decreases in heart rate variability (Ottaviani, Shapiro, & Couyoumdjian, 2013), and increases in blood pressure (Ottaviani, Shapiro, & Fitzgerald, 2011). Hence, sticky thoughts may not only be disruptive to task performance, but also pose a risk for developing mental and somatic health issues.

3.1.2 Examining stickiness of thought with self-report and task performance

Despite the known disruptive effects of sticky thought on task performance, we have limited understanding of the (attentional) processes that are associated with sticky thought and how those differ from non-sticky thought. One reason for this is that sticky thought is challenging to detect in the context of an experiment (Smallwood & Schooler, 2015). Sticky thinking is largely a covert process that leaves few directly observable signs. Indeed, related processes such as perseverative cognition and rumination have mostly been investigated using self-report questionnaires that measure trait rumination or worry (i.e., the general tendency to engage in sticky thinking), or alternatively, by asking the participant to report on the frequency of ruminative or worry episodes retrospectively. Correlating such measures with task performance (Beckwé & Deroost, 2016; Joormann et al., 2011; Koster, De Lissnyder, & De Raedt, 2013) and neurocognitive measures (e.g., Ottaviani et al., 2013) has yielded valuable insights (see Susan Nolen-Hoeksema et al., 2008). For example, Beckwé et al. (2016) found that in an exogenous cue task (ECT) participants with a strong tendency to ruminate had longer reaction times following invalid negative personality trait cues, suggesting that such participants experience more difficulty to disengage from negative personality trait cues, likely because these cues set off a train of negative self-related thinking. Aside from this cognitive inflexibility, studies with cardiac measures have shown that rumination is associated autonomic rigidity, demonstrated by persistent low heart rate variability (Ottaviani et al., 2013, 2011). Despite these insights from questionnaire-based measures of sticky thoughts, self-report arguably lacks precision, given limitations in memory that bias reporting (Schwarz, 2007; Scollon, Kim-Prieto, & Diener, 2003) and participants' tendency to produce socially desirable answers. Furthermore, because questionnaires only provide a single after-the-fact measure, it is not possible to compare sticky with non-sticky thought within an individual.

A different, and potentially better, method to measure sticky thought are thought probes. Thought probes are short self-report questionnaires that are embedded in a task to measure the content and dynamics of current thought at various points in time

during an on-going task (Kane et al., 2017; Weinstein, 2018). They have the advantage that experiences can be caught close to when they arise. Furthermore, they allow for repeated measures of experienced thought making it possible to investigate changes in thought content over the course of the experiment. For example, Unsworth and Robison (2016) used thought probes to investigate how different attentional states, such as mind wandering and external distraction, correlated with task performance and pupil size measures in sustained attention task. The researchers observed that task performance decreased and pupil size became smaller with time-on-task. Also, they found that reports of mind wandering were more frequent when the experiment progressed. This demonstrates that time-on-task influences are important to consider when studying self-generated thinking.

So far, we are familiar with only one study that used thought probes to investigate sticky thought. Van Vugt and Broers (2016) used thought probe responses in conjunction with task performance measures to investigate how self-reported stickiness of thought was associated with the probability of being disengaged from the task (i.e., off-task). In addition, the researchers examined how self-generated thought and its stickiness affected performance. They asked participants to perform a variation of a go/no-go task referred to as the sustained attention to response task (SART; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). This task is suitable for studying self-generated thought because it is slow-paced and induces habitual responding, therefore allowing self-generated thought to occur. In line with their expectation, self-reported stickiness of thought increased the probability of being disengaged from the task and negatively influenced performance. Stickiness of thought was associated with more variable response times. Previous research has also indicated that variability in response times may be a relevant correlate for self-generated thought (see e.g., Bastian & Sackur, 2013; McVay & Kane, 2012; Seli et al., 2013). The increase in variability may indicate that participants allocate less attention to the task, resulting in reactive and more variable responding (Braver, 2012). All in all, this study demonstrated that stickiness of thought is a relevant dimension of self-generated thought. Furthermore, it indicated that people can meaningfully report on the stickiness of their thought.

3.1.3 Correlates of self-generated thought and stickiness in pupillometry

In addition to task performance, neurocognitive measures can be used to detect sticky thought, and can provide insight into the processes and mechanisms associated with sticky thought. In this study, we will use pupillometry to gain insight in sticky thought. Pupil size is an interesting measure because it relatively unobtrusive and easy to record. Furthermore, research has indicated that lapses of attention can be distinguished in various pupil size measures (Huijser, van Vugt, & Taatgen, 2018; Unsworth & Robison,

2016). Therefore, we may also be able to detect differences in pupil size depending on the stickiness of thought.

Pupil size is typically measured on two temporal scales, reflecting different cognitive or neural processes. The most common of these measures is the task-evoked response in pupil size. The task-evoked response is a transient increase in pupil size following the processing of a task event, peaking at around 1s after event onset (Hoeks & Levelt, 1993). The magnitude of this response has been demonstrated to depend on the amount of attention, cognitive control, and cognitive processing required by the task (Gilzenrat et al., 2010; Unsworth & Robison, 2014; Wierda et al., 2012). Research has consistently found that when we engage in self-generated thought, our evoked responses in pupil size are smaller (Grandchamp et al., 2014; Jubera-García, Gevers, & Opstal, 2019; Mittner et al., 2014; Unsworth & Robison, 2016). Smallwood and colleagues (Smallwood, Brown, et al., 2011) interpreted this smaller response in pupil size as evidence that external processing is being inhibited during self-generated thinking, so-called perceptual decoupling.

In addition to stimulus-evoked pupil responses, pupil size is also measured during task-free periods, referred to as baseline or tonic pupil size. Baseline pupil size is proposed to reflect locus coeruleus norepinephrine (LC-NE) system functioning, which has been associated with controlling overall arousal levels and the tendency to seek for novelty (Murphy et al., 2014, 2011; Unsworth & Robison, 2016). Large baseline pupil size has been correlated with high tonic LC-NE firing, indicating a state of over-arousal and tendency to explore new behaviors. On the other hand, smaller baseline pupil sizes have been related to low tonic firing, under-arousal, and inactivity. Interestingly, research has proposed that the relationship between baseline pupil size, task-evoked pupil size, and task performance can be described with an adaptive gain curve (see Figure 3.1; see e.g., Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010). Task performance is optimal at intermediate levels of baseline pupil size when task-evoked responses are maximal. Task performance decreases when baseline pupil size is either larger or smaller.

Since stickiness (i.e., the difficulty in disengaging from thought) is a novel topic, no studies have directly investigated how stickiness is reflected in baseline and task-evoked pupil size. Nonetheless, predictions can be made based on related research. Given the disruptiveness of sticky thought to ongoing activities, we may expect that sticky thought, similar to self-generated thinking, is associated with smaller task-evoked responses in pupil size. As predicted by adaptive gain (see Figure 3.1), a smaller task-evoked response in pupil size with episodes of sticky thought would imply that the thought process is associated with either smaller or larger than average baseline pupil size. However, which one is open to debate. In clinical samples, Siegle et al. (2003) found that rumination

Figure 3.1. The adaptive gain curve describes the relationship between baseline pupil size and task-evoked pupil size (Aston-Jones & Cohen, 2005; Murphy et al., 2011). Task-evoked responses are maximized at intermediate levels of baseline pupil size, but decrease in magnitude when the baseline is smaller or larger. The curve also makes predictions about task performance. Performance on a task is optimal at intermediate baseline pupil size when task-evoked responses are maximal. Task performance decreases when baseline pupil size is smaller or larger than the intermediate level.

was associated with larger baseline pupil sizes. The researchers hypothesized that this larger baseline pupil size reflected sustained emotional processing (Duque, Sanchez, & Vazquez, 2014). In contrast, Konishi et al. (2017) found that in non-clinical samples, negative and intrusive thoughts were associated with smaller baseline pupil size (Harrison, Singer, Rotshtein, Dolan, & Critchley, 2006). One recent study investigated how the “intensity” of experienced thought was reflected in baseline pupil size, which may be a dimension somewhat comparable to the stickiness of the thought (Jubera-García et al., 2019). However, that study was unsuccessful in finding an effect of intensity of thought in baseline pupil size. Also for self-generated thought, the literature has not reached consensus on where the thought process lies on the adaptive curve (Franklin et al., 2013; Grandchamp et al., 2014; Smallwood, Brown, et al., 2011). Given that sticky thought is proposed to develop from thinking about pressing concerns and unreach goals (Klinger, 2013; Marchetti et al., 2016), one might think that sticky thought is associated with high arousal, and therefore larger than average baseline pupil size. On the other hand, one might also think that sticky thought results from a state of inertia, reflected in smaller baseline pupil size.

3.1.3 The current study

The main goal of the present study was to investigate how stickiness of thought is reflected in task performance and pupillary measures (i.e., baseline and task-evoked

response in pupil size). We asked participants to perform a variation of the SART, in which we embedded the personal concerns of participants in the task to potentially increase the tendency for sticky self-generated thought (see McVay & Kane, 2013). We included periodic thought probes in the SART to measure what participants were currently thinking about (i.e., attentional state) and how difficult it was to disengage from the thought (i.e., stickiness of thought).

In line with previous work, we expected that sticky thought would be associated with being more disengaged from the task. Since being disengaged from the task has been found to reduce no-go accuracy, speed-up response times and increase RTCV, we predicted that no-go accuracy would decrease, response times would be faster, and that RTCV would increase with reported stickiness of thought.

With respect to pupillary measures, we predicted that stickiness of thought would be associated with smaller task-evoked responses in pupil size, indicating reduced attention to the SART. Given that no research has investigated the influence of stickiness of thought on baseline pupil size, and the inconsistency in previous studies that tried to relate for baseline pupil size to self-generated thought, we formulated no prior hypotheses for that measure.

3.2 MATERIALS AND METHODS

3.2.1 Participants

We recruited 34 Native Dutch speakers for this experiment (20 female; $M_{\text{age}} = 22.7$, $SD_{\text{age}} = 2.7$). Participants were recruited from a paid research participant pool on Facebook, as well as from the Artificial Intelligence Bachelor and Master programs at the University of Groningen. We screened the participants for having normal or corrected-to-normal vision prior to testing. All participants provided informed consent at the start of the laboratory session. The experiment was conducted in accordance with the Declaration of Helsinki and approved by the Ethical Committee of Psychology (ECP) at the University of Groningen (research code: pop-015-170). Written informed consent was obtained prior to the experiment for each participant.

3.2.2 Materials

3.2.2.1 Questionnaire Session

Participants were requested to fill out three online questionnaires prior to the experiment. Since we wanted to maximize the probability of observing sticky thinking, which is known to often be related to concerns and worries, we adopted the current concerns manipulation by McVay & Kane (2013). For this manipulation, individual current concerns were collected using an online version of the Personal Concerns

Inventory (PCI; adapted from Klinger & Cox, 2004). In this questionnaire, participants were asked to write down short statements about current goals or concerns in eight different areas, including: 1) home and household matters, 2) employment and finances, 3) partner, family, and relatives, 3) friends and acquaintances, 4) spiritual matters, 5) personality matters, 6) education and training, 7) health and medical matters, and 8) hobbies, leisure, and recreation. For every current goal or concern, participants were asked to rate the importance on a scale from one to ten, and to indicate a time frame in which the goal/concern was expected to be accomplished or resolved. Participants were encouraged to think about goals or concerns that were relevant in the coming year. In addition to the PCI, the Behavioral Inhibition System/Behavioral Approach System scales (BIS/BAS; Carver & White, 1994) and the Habit Index of Negative Thinking (HINT; Verplanken, Friborg, Wang, Trafimow, & Woolf, 2007) were used as distractor questionnaires to make the goal of our study less obvious to our participants. The PCI and BIS/BAS questionnaires were administered in Dutch (translated from English), the HINT in English (original language) given that no validated translation was available.

3.2.2.2 *Experimental Session*

The SART in this experiment was based on the task used by van Vugt and Broers (2016) and McVay and Kane (2013). Our SART included 720 Dutch words as stimuli that were presented in black. The majority of words were lower-case go stimuli ($n = 640$, 89% of total set), while only a small set were upper-case no-go stimuli ($n = 80$; 11% of total set). Participants were instructed to press a button as fast as they could on go stimuli, but to withhold a response on the infrequent no-go stimuli. All stimuli were presented centrally against a grey background.

Similar to the earlier works we embedded participant's personal current concerns in the SART task, along with the current concerns from another participant as a control (i.e., other concerns). We selected two personal concerns for each participant based on the PCI answers, and two 'other' concerns that were distinctly different from their personal concerns. Each current concern was translated into a triplet of words. For example, if a participant reported (A), this was translated into (B).

A. "Er zijn nog wat dingen die ik moet voorbereiden voordat ik kan beginnen met een tussenjaar." (English: "There are still some things I need to arrange before I can start taking a gap year.")

B. pauze loopbaan prepareren (English: break career preparation)

We looked for two personal concerns with the highest importance rating. Whenever

two concerns had the same importance rating, we selected the most unique concern. Concerns that were too common or general were avoided. Concern words were always go stimuli. The stimulus words that were not part of the personal/other concern triplets were selected from the Dutch word frequency database: SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010). This database contains word frequency values based on film and television subtitles. We selected the stimulus words based on the *Lg10CD* variable. This variable is a measure of the contextual diversity of a word, reflected in how many films or television shows it occurred. A validation study with a lexical decision task showed that the *Lg10CD* variable explained most variance in task performance (i.e., accuracy and response time; see Keuleers et al., 2010). The same study also showed that the SUBTLEX-NL database explains 10 % more variance compared to the common CELEX database (Baayen, Piepenbrock, & van Rijn, 1993). Before selecting the word stimuli, we first discarded the least and most frequent words from the database. Thereafter, 312 words were selected with a *Lg10CD* value around the mean. We removed and replaced selected stimuli that were numbers, non-words, or high-arousal words.

We measured the occurrence of self-generated thought and the stickiness of thought by periodically including thought probes in the task. Thought probes consisted of two questions (Figure 3.2). The first question was adopted from Unsworth and Robison (2016) and addressed the current thought content or attentional state. This question differentiated six types of attentional state: 1) on-task focus, 2) task-related interference (TRI), 3) concern related thought, 4) external distraction, 5) mind wandering, and 6) mind blanking/inattentiveness. The second question was adopted from van Vugt and Broers (2016) and asked how “sticky” the current thoughts were. Stickiness was measured as thought being 1) very sticky, 2) sticky, 3) neutral, 4) non-sticky, and 5) very non-sticky. We included 48 thought probes in the experiment.

It is relevant to note that the second ‘stickiness’ question has only been used once in previous research (see van Vugt & Broers, 2016), while the question on attentional state (and similar counterparts) have been used more frequently. Therefore, the reliability and validity of the measure cannot be guaranteed. However, there are signals that provide confidence in the reliability and validity of the stickiness question. First, the significant differences in task performance across the different levels of stickiness reported by the study of van Vugt and Broers does indicate that participants are able to report on the stickiness of their thought with similar accuracy to other thought responses. Furthermore, Mills et al. (2018) showed that participants’ assessment of the extent to which their thoughts were constrained, a concept similar to our stickiness, correlated significantly with external reviewers’ assessments.

<p>What were you just thinking about?</p> <ol style="list-style-type: none"> 1) I was completely focused on the task. 2) I was evaluating aspects of the task. (For example my performance or how long it is taking) 3) I was thinking about personal matters. 4) I was distracted by my environment. (For example sound, temperature, my physical state) 5) I was daydreaming/ I was thinking of task-unrelated matters. 6) I was not paying attention, but neither was I thinking about anything specifically. <p>Press the number that matches your answer to continue</p>	<p>How difficult was it to disengage from the thought?</p> <ol style="list-style-type: none"> 1) Very difficult 2) Difficult 3) Neither difficult nor easy 4) Easy 5) Very easy <p>Press the number that matches your answer to continue</p>
---	---

Figure 3.2. Thought probe question. An English translation of the thought probe questions used in the experiment. The first (left) question was used to measure attentional state, the second (right) question to measure stickiness of thought.

3.2.3 Apparatus and set-up

Participants completed the PCI, HINT, and BIS/BAS questionnaires online, prior to coming to the lab, using Google Forms. The SART was performed individually in the lab. This lab contained a desk for the participants on which a computer, monitor, eye tracker, and head-mount was located. Pupil size and gaze position of the dominant eye were recorded at a sampling rate of 250 Hz using an Eyelink 1000 eye tracker from SR Research. The experiment was programmed in Psychopy (version 1.83.04; Peirce et al., 2019) and interfaced on a Mac mini running Windows 7. The stimuli were presented on a 20 inch LCD monitor with a resolution of 1600x1200 pixels (4:3 aspect ratio) and a refresh rate of 60 Hz.

3.2.4 Procedure

3.2.4.1 Questionnaire Session

Following registration for the experiment, participants received an email with a single link to the three online questionnaires. Participants started with the HINT, followed by the BIS/BAS and PCI questionnaire respectively. They were instructed to complete the questionnaires no later than the day before the laboratory session. After filling out the questionnaires and before the experimental session, we collected the current concerns from the answers on the PCI as described in section *Materials: Experimental Session*. The selected concerns, together with the concerns of another participant, were subsequently

Figure 3.3. Task overview. A series of trials (top) and a single trial (bottom). Whenever a concern triplet was presented (red boxes), this was followed by four go trials (blue boxes), one no-go trial (green box), and one thought probe. Each trial, go or no-go, started with a variable inter-trial interval (ITI). Thereafter, the stimulus was presented in lowercase for go trials and uppercase for no-go trials. The stimulus was followed by a mask (until response) and a response interval (until response). Whenever the participant responded during the mask or the response interval, the experiment immediately proceeded with the next trial (i.e., ITI is drawn).

embedded in the stimulus set of the respective participant.

3.2.4.2 Experimental Session

The experimental session started with setting up the eye tracker. Participants were seated in front of the display computer and monitor, eye tracker, and head-rest. The head-rest was adjusted to the height of the participant. We performed a nine-point calibration and separate validation using the eye tracker software. The calibration and validation procedure were performed for the dominant eye of the participant, or in some cases the other eye if that provided a better signal. Following calibration and validation, the instructions for the experiment were presented on the screen. The instructions on how to perform the SART were presented first, including one example of a go and a no-go trail. Afterwards, participants were informed that they would be periodically asked to report on their current thoughts. The questions for attentional state and stickiness of thought were presented on the screen, including the instructions on how to report their answer. The participants were not otherwise instructed or trained on how to use the thought probes but were invited to ask questions any time. A short practice session fol-

lowed the instruction phase. This practice session consisted of ten SART trials (including one no-go trial) and one thought probe. The practice session included no trials reflecting a *personal* or *other* concern. After practice, the experiment started and the eye tracker started recording.

Each trial (see Figure 3.3, bottom) started with an inter-trial interval (ITI) of variable duration between 1500 and 2100 ms. During the ITI, a fixation cross consisting of the ‘+’ symbol was presented centrally on the screen. The ITI was followed by the presentation of the stimulus word for 300 ms. Go stimuli were presented in lower-case, whereas no-go stimuli were presented in upper-case. Participants were instructed to only respond on go trials (as fast as possible) by pressing the ‘m’ key on the keyboard and to withhold a response on no-go trials. After stimulus presentation, a mask (‘XXXXXXXX’) was presented for 300 ms followed by a response interval of 3000 ms marked by a ‘+’ symbol. Pupil responses were recorded during the stimulus, mask, and response intervals. Once a participant responded during the mask or response interval, the experiment immediately moved on to the next ITI.

The 720 trials in the experiment (640 go; 80 no-go) and 48 thought probes were equally distributed across eight blocks of 90 trials (80 go; 10 no-go) and six thought probes. All participants saw the same (no concern) stimulus words but in a random order. The blocks consisted of two similar sequences of 45 stimulus words and three thought probes. The only difference between the two sequences in a block was the concern condition. One sequence contained a *personal* concern triplet, whereas the other contained an *other* concern triplet. The order was counterbalanced across the experiment. Furthermore, each block contained only one of the two *personal* and *other* concerns. Which type of concerns was selected alternated between blocks. When a concern triplet was presented, the order of the trial type (i.e., go – personal concern, go – other concern, go – no concern, no-go) and thought probes was fixed. This order was based on the experiment of McVay and Kane (2013). As shown in Figure 3.3 (top), concern triplets were always followed by four go (no concern) trials, one no-go trial, and one thought probe. The thought probe questions always immediately followed the no-go trial to ensure that the reported thought content and its stickiness could be reliably attributed to the trials before it. We are aware that a limitation of this design is that participants may confabulate their answer to the thought probe as being off-task when an error has been made on the no-go trial. Nevertheless, since this is the procedure used many prior studies on which we based our work, we kept this design.

3.2.5 Data analysis

3.2.5.1 *Preprocessing of eye tracking recordings*

Before analysis, we first removed pupil size measurements associated with blinks and other artifacts. Blinks were detected using the eye tracker software. We removed the pupil size measurements marked as a blink including 100 ms before and after the event. In addition, we removed sudden upward or downward jumps. Jumps were identified by first z-scoring the pupil size timeseries for each participant individually. Subsequently, we marked pupil size measurements that had a 0.05 absolute difference in pupil size from the previous measurement (i.e., 4 ms earlier with a 250 Hz sample rate) including 20 ms before and after the observation. Subsequently, we visually inspected the marked segments of the data that would be removed with this cut-off. We concluded that this cut-off was sensitive enough to remove the jumps, but not so sensitive that it would also discard ‘normal’ increases in pupil dilation. In total we discarded 12.2 % of the pupil size measurements with SART trials, with percentages ranging from 1.7 % to 29.4 % across individual participants. Trials with more than 25 % discarded/missing data were removed completely, resulting in the removal of 11.1 % of the trials (range = 0.3 % - 57.3 %). We downsampled the data to 50 Hz, taking the median pupil size for each time bin. We did not interpolate the data, since our analysis methods (generalized additive mixed models and linear-mixed effect models; see *Statistical Analysis*) can deal with missing data. After downsampling, we segmented the pupil size measurements in timeseries for individual trials ranging from 500 ms before stimulus onset to 2000 ms after onset. This time window was chosen to fully capture the pupil response to the task stimuli, while preventing overlap in the segments of neighboring trials.

3.2.5.2 *Eye tracking measures*

We calculated the baseline pupil size by first taking the mean of the pupil size measurements in the window of 500 ms before stimulus onset. Subsequently, the baseline pupil size was then determined by z-scoring the means for each participant individually. Z-scoring the baseline values sets the grand average for each participant at zero, thereby removing individual differences in pupil size. Task evoked pupil size was obtained by subtracting the (non-transformed) baseline of each trial from the pupil size measurements in that trial.

3.2.5.3 *Behavioral measures*

We measured task performance in the SART on accuracy, response time, and variability in response time (RTCV). Accuracy was expressed as a binomial dependent variable, coding correct responses as ‘1’ (i.e., button press on go trials and no response on no-go trials) and incorrect responses as ‘0’ (i.e., no response on go trials and a button press on no-go trials). Response was measured in milliseconds, but log-transformed to account

Answer option	Frequency (out of N = 48 per subject)	Percentage
On-task	22.3	46.5
Task-related interference	10.5	21.8
Current concerns	4.9	10.1
External distraction	5.2	10.8
Mind wandering	3.9	8.0
Inalertness	1.3	2.3

Table 3.1. Distribution of responses to attentional state question. Average number of responses (out of N = 48) to each answer option on the attentional state question per subject. Relative frequencies, expressed in percentages, are presented in the third column.

Answer option	Frequency (out of N = 48 per subject)	Percentage
Very sticky	3.4	7.4
Sticky	12.4	25.9
Neutral	19.8	41.1
Non-sticky	7.9	16.5
Very non-sticky	4.5	9.4

Table 3.2. Distribution of responses to stickiness question. Average number of responses (out of N = 48) to each answer option on the stickiness question per subject. Relative frequencies, expressed in percentages, are presented in the third column.

for right-skewness in the distribution of these measurements. RTCV was calculated by taking the mean of the four go trials preceding a no-go trial (Figure 3.3), divided by the standard deviation of response time in those trials. Similar to the response times, the RTCV values were log-transformed prior to analysis.

Alongside these task performance measures, we included variables for the current concerns condition, attentional state (i.e., on-task, self-generated thought, task-related interference etc.), and stickiness level (i.e., very non-sticky, non-sticky, neutral, sticky, very sticky) in the analysis. Stickiness level was both used as (ordered) categorical dependent variable and predictor depending on the analysis. Current concerns condition and attentional state were only included as categorical predictors. The current concern condition predictor indicated whether a go/no-go trial was preceded by a triplet of

personal or other concerns, or in cases where there were no concern related trials, within a window of eight trials (see Figure 3.3). The attentional state and stickiness level predictor indicated the answer on the first and second thought probe question respectively.

We noticed that some answer options on both thought probe questions had very few observations (see Table 3.1 and 3.2). To increase the amount of observations per answer option and thereby increase statistical power, we decided to combine the answer options into larger categories. For attentional state, we combined the answer option for thoughts about current concerns (option 3) with mind wandering (option 5) into the larger category of *self-generated thought*, justified by the idea that thoughts about concerns are a special case of mind wandering. We also combined the option for external distraction (option 4) and inattention (option 6) into the *other* category. This resulted in the following levels: on-task, task-related interference, self-generated thought, and other. For stickiness level, we decided to group the first two answer options into a *sticky* category, and the last two into *non-sticky*. We refer to the third (intermediate) answer option as *neutral*.

3.2.5.4 Statistical analysis

We investigated the thought probe reports by computing a count for each answer option to both questions for each participant. The resulting answer frequencies were analyzed using generalized linear models assuming a Poisson distribution. We used linear-mixed effects modeling (LME) in the remaining analyses, except when ‘time’ (i.e., time-in-trial, time-on-task) was considered as a predictor. We assumed a Gaussian distribution for fitting response time, RTCV, and baseline pupil size. A binomial distribution was assumed for accuracy measurements. We fitted an ordered categorical LME for predicting stickiness level.

When fitting LMEs, it is important to determine a good random effects structure. Including too few random effects makes the model potentially over-confident, resulting in more Type I errors (Barr, Levy, Scheepers, & Tily, 2013). Including too many random effects lowers statistical power (Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). To balance Type 1 error and statistical power, we determined the random effects structure of each model using a chi-square log-likelihood-based backwards model fitting procedure. With this procedure we removed one term from the random effects structure at every step, starting from the most complex model. We kept the simpler model if the more complex model did not significantly explain more variance. Random effects that correlated strongly ($r > 0.5$) with one or more other random effects were always removed. Models that did not converge or provided a singular fit

were not considered. In such cases, we continued the procedure of leaving one term out at every step. We considered trial number, block number, and participant number as random intercepts. Current concern condition, attentional state, and stickiness level were considered to have random slopes whenever they were included as a fixed effect in the model.

Statistical significance of individual predictors in the fitted LMEs were determined using chi-square log-likelihood ratio tests, testing the model including the predictor against an intercept-only model. Interactions were tested by comparing a model with the interaction against a model with only the main effects. Predictors in the LMEs were categorical. Consequently, the test statistics only reflect comparisons to a reference group of the categorical predictor(s). The reference group for attentional state was 'on-task', for stickiness of thought 'neutral', and for current concerns the 'no concern' condition. Regression estimates (i.e., intercept and slopes) of individual LMEs were transformed back to the original scale to enhance interpretation. For Gaussian LMEs we did not determine p-values, but we use report t-statistics to indicate statistical significance ($|t| \geq 2$).

We conducted timeseries analysis (e.g., for task-evoked pupil response; time-on-task effects on attentional state, stickiness, and baseline pupil size) using a nonlinear regression technique called generalized additive mixed modeling (GAMM; Wood, 2011, 2017a). Unlike existing related research using summarizing measures such as mean pupil size after stimulus onset (Grandchamp et al., 2014), the slope of pupil size (Grandchamp et al., 2014; Mittner et al., 2014; Mittner, Hawkins, Boekel, & Forstmann, 2016), and the maximum pupil size in a specified time window (Unsworth & Robison, 2016), GAMM allows you to model full time courses. The difference from linear regression is that the slope estimates are replaced by smooth functions that describe how a timeseries measure such as task-evoked pupil size changes over time. When a categorical predictor is added to the GAMM, the model will fit a different smooth function for every level of this predictor. Such smooth functions can subsequently be visualized to examine the development of the statistical effects over time. GAMM also allows for including nonlinear random effects called random smooths. In essence, random smooths estimate random effects coefficients for the intercept as well as how the slope of a timeseries changes over time. In our analyses, we used a random smooth for events that reflected the individual time course of each trial and participant. Alongside a random smooth for events, we also included a nonlinear interaction between the x and y gaze position in each GAMM. This was to account for influences of gaze position on pupil size (Brisson et al., 2013; Gagl, Hawelka, & Hutzler, 2011). An issue with modeling task-evoked responses in pupil size is that the residuals of the model are not normally distributed. To account

for non-normality, we fitted all GAMMs for task-evoked pupil responses (except for one) assuming a scaled-t distribution (Wood, Pya, & Saefken, 2016). Only the GAMM estimating the influence of stickiness on go-trial evoked responses was fitted as a Gaussian model, since the model did not converge when assuming a scaled-t distribution. Another common issue is that the pupil size recordings are highly correlated over time, violating the method's assumption that residuals are independent. Violation of this assumption may cause a GAMM model to underestimate the size of standard errors. We accounted for autocorrelation by including an autoregressive AR(1) error model within each GAMM (Wood, Goude, & Shaw, 2015). For an excellent tutorial paper on how to use GAMMs for pupil size analysis, we refer to van Rij and colleagues (2019).

From the fitted GAMMs, we could determine whether estimated smooth terms were statistically significant. In other words, we could determine whether there were significant (nonlinear) changes in the value of a dependent variable (such as task-evoked pupil size) along the time course of a trial for different attentional states and stickiness levels. We checked for significant differences between two timeseries by determining a difference curve based on the estimated smooth terms. Two timeseries were considered to be significantly different at some point in time when the estimated difference curve including a pointwise 95% confidence interval did not include zero (given that zero indicates the absence of a difference).

Preprocessing and data analysis were performed in R. We used the lme4 package to fit the Gaussian and binomial LMEs (version 1.1-19; Bates, Maechler, Bolker, & Walker, 2018). GAMMs and ordered categorical LMEs were fitted using the mgcv package (version 1.8.28; Wood, 2017b). Model estimates and diagnostics for GAMMs were visualized with the itsadug package (version 2.3; van Rij, Wieling, Baayen, & Van Rijn, 2017). The data and the analysis code for preprocessing and model fitting are available online at: <https://osf.io/m6ujg/>.

3.3 RESULTS

3.3.1 Thought reports

First, we analyzed the reports collected from the two thought probe questions. We assessed whether embedding participant's personal concerns influenced participants' tendency to engage in sticky or off-task thinking. Next, we examined whether time-on-task influenced attentional state (answer to the question "what were you just thinking about?") and stickiness of thought (answer to the question "how difficult was it to disengage from the thought?"). Finally, we investigated the relationship between attentional state and the experienced stickiness of thought.

Figure 3.4. Influence of concern manipulation on frequency of different attentional states. Report counts were derived from the first thought probe question. This question had six answer options. From left to right on the x-axis, on-task refers to the answer option for task focus, TRI to task-related interference, SGT (i.e., self-generated thought) to answers on mind wandering and thoughts on personal concerns, and other to answers on external distraction and inattention. Error bars reflect one standard error of the subject mean.

3.3.1.1 Current concerns manipulation

Figure 3.4 shows the effect of the current concerns manipulation. Following no concern triplets, we observed that on-task reports were most frequent ($M = 7.97$ (in count) out of $N = 48$ total reports; $SD = 3.34$), followed by reports of task-related interference ($M = 4.00$; $SD = 2.67$), self-generated thought ($M = 2.06$; $SD = 1.76$), and other reports ($M = 1.97$; $SD = 1.85$). The number of self-generated thought reports increased after a personal concern triplet relative to a no concern triplet ($M_{diff} = +1.59$). The average increase in self-generated thought reports after concerns was significant ($\beta = +1.59$ (in count), $z = -2.12$, $p < .001$). In addition, we found that concerns from another participant increased the amount of self-generated thought reports ($M_{diff} = +0.94$; $\beta = +0.94$ (in count), $z = -2.73$, $p = .03$). However, the mean increase in frequency of self-generated thought following such “other concerns” was found to be smaller compared to personal concerns ($M_{diff} = -0.65$; $\beta = -0.65$ (in count), $z = -2.52$, $p = .009$). Therefore, while personal concerns and other concerns were both found to increase self-generated thinking, personal concerns were more potent.

With respect to the stickiness of thought, we found that participants most frequently reported their thought as neutral following no concern triplets ($M = 6.71$ (out of 48 total

Figure 3.5. Observed and estimated effect of time-on-task on attentional state. The left plot presents the observed data and the right plot presents the estimated data from the best-fitting GAMM model. The GAMM model explains 27% of the variance in the thought probe data, calculated by taking the square of the correlation between the observed (fitted) and predicted data. Error bars in the right plot reflect estimated pointwise 95% confidence intervals.

reports); SD = 4.48), followed by sticky (M = 5.06; SD = 3.85), and non-sticky (M = 4.24; SD = 4.36). In contrast to what we found for self-generated thought, we did not find support for an increase in stickiness of thought following personal (or other) concerns ($\chi^2(2) = 3.82, p = .15$). Therefore, it is unclear whether processing current concerns in the SART could increase the stickiness of thought.

3.3.1.2 Time-on-task influence

Figure 3.5 shows how attentional state (e.g., on-task, self-generated thought etc.) changed over the course of the experiment. The right figure shows the estimates from the fitted GAMM. Our results indicated that only the smooth terms for on-task and self-generated thought were significant (on-task: $F = 8.00, p = .005$; self-generated thought: $F = 10.66; p = .001$; task-related interference: $F = 2.90, p = 0.09$; other: $F = 1.33, p = 0.31$). Therefore, we can (only) conclude for on-task and self-generated thought that the amount of reports on this type of thinking changed over time. As shown in Figure 3.5, on-task thought decreased while self-generated thought increased as the task progressed.

We then asked how stickiness of thought changed over the course of the experiment. We fitted an ordered-categorical GAMM to test how time-on-task influenced the likelihood of reporting having neutral, sticky, or non-sticky thoughts. For this analysis we included

Figure 3.6. Observed and estimated effect of time-on-task on stickiness level. The left plot presents the observed data and the right the estimated data from the best-fitting GAMM model. The estimated probabilities in the right plot were derived by fitting an ordered-categorical GAMM model. Error bars in the right plot reflect estimated pointwise 95% confidence intervals.

the reported answer options as an ordinal dependent variable (1 being non-sticky, 2 neutral, and 3 being sticky). Block number was included as continuous predictor reflecting time-on-task. The results showed that the smooth term for block number was significant ($\chi^2 = 12.11$, $p < .001$), indicating that the reported level of stickiness changed over the course of the experiment. To inspect how the likelihood of reporting the different levels of stickiness changed over time, we obtained the predicted probability estimates from the model and plotted these in Figure 3.6 (right). With increasing time-on-task, we found that reports of neutral thought remained relatively constant. Furthermore, neutral thought was most prevalent in general. At the same time, we found that the probability of sticky thought increased with time-on-task, while it decreased for non-sticky thought. Together, this indicates that thought became more sticky as the task progressed.

3.3.1.3 Relationship between attentional state and stickiness level

We then examined whether the level of stickiness depended on whether a participant was focused on the task, mind-wandering or elsewhere. As shown in Figure 3.7, the reported stickiness of on-task thought strongly differed from the distracted attentional states. The majority of non-sticky ($M = 0.58$) and neutral thought reports ($M = 0.61$) were associated with on-task thought. On the other hand, reports of sticky thought were relatively more frequent in the distracted states. To test whether distracted states

Figure 3.7. Relationship between stickiness of thought and attentional state. Report counts were derived from the first (attentional state) and second thought probe question (stickiness of current attentional state). Error bars reflect one standard error of the subject mean.

were experienced as stickier, we fitted an ordered categorical (ordinal) LME predicting stickiness level by attentional state. The model indicated that all off-task states were reported as stickier than on-task (on-task: intercept $\beta = -0.40$ (transformed), $t = -1.79$; self-generated thought: $\beta = +1.53$ (transformed), $t = 9.85$; task-related interference: $\beta = +1.54$ (transformed), $t = 11.07$; other: $\beta = +1.77$ (transformed), $t = 10.11$).

3.3.2 Task performance

We analyzed task performance to examine how attentional state and stickiness of thought was reflected in performance on go and no-go trials. Overall, we found that participants were 56.62 % accurate (SD = 49.57 %) on no-go trials. The mean response time to go trials was 375.51 ms (SD = 94.14 ms), with a mean coefficient of variance (RTCV) of 0.14 (SD = 0.12). As expected, we found that all ‘distracted’ attentional states were associated with a lower accuracy on no-go trials compared to on-task ($\chi^2(3) = 216.08$, $p < .001$; on-task: intercept $\beta = 0.80$, $z = 5.92$, $p < .001$; self-generated thought: $\beta = -0.36$, $z = -9.47$, $p < .001$; task-related interference: $\beta = -0.$, $z = -11.54$, $p < .001$; other: $\beta = -0.42$, $z = -10.08$, $p < .001$). No significant influence of attentional state was found on go response time ($\chi^2(3) = 4.88$, $p = .18$) nor on RTCV ($\chi^2(3) = 5.43$, $p = .14$). For stickiness of thought (see Figure 3.8), the results showed neither a significant influence of stickiness on response time ($\chi^2(2) = 2.37$, $p = .31$), nor on RTCV ($\chi^2(2) = 2.68$, $p = .26$). On the other

Figure 3.8. Influence of stickiness on task performance. Mean no-go accuracy (left), go response time (center), and go RTCV (right) for each level on the stickiness dimension. Error bars reflect one standard error from the subject mean.

hand, we did find a significant step-wise decrease in no-go accuracy from non-sticky thought to sticky thought. Compared to neutral stickiness, participants were 20 % more accurate when current thinking was non-sticky ($\beta = +0.20$, $z = 5.96$, $p < .001$), and 29 % less accurate when current thinking was more sticky than neutral ($\beta = -0.29$, $z = -8.12$, $p < .001$). When attentional state was added to the LME model as an additional categorical factor alongside stickiness, we found that stickiness remained a significant predictor of no-go accuracy ($\chi^2(2) = 82.99$, $p < .001$), but not RT ($\chi^2(2) = 1.14$, $p = .57$) or RTCV ($\chi^2(2) = 0.83$, $p = .66$). This suggests that stickiness exerts unique influence on no-go accuracy on top of attentional state. The model predicted that participants were 23% more accurate when self-generated thinking was non-sticky ($\beta = +0.23$, $z = 5.35$, $p < .001$) compared to neutral (intercept $\beta = 0.48$, $z = -0.25$, $p = .80$). Participants were 17 % less accurate when self-generated thought was reported as sticky ($\beta = -0.17$, $z = -4.63$, $p < .001$).

3.3.3 Baseline pupil size

The behavioral results indicated that the frequency of different attentional states and stickiness of thought changed over the course of the experiment. Therefore, we need to take time-on-task into account when we assess how attentional state and stickiness are reflected in baseline pupil size. Figure 3.9 (top panel) shows the baseline pupil size across blocks for each attentional in the data (left) and predicted by a GAMM model (right). The data and the model demonstrated that baseline pupil size became smaller as the task progressed. At the same time, we failed to find consistent differences in the baseline pupil size between the attentional states. Therefore, we cannot conclude that

Figure 3.9. Observed and estimated baseline pupil size for different attentional states and stickiness levels. The left plots show the observed baseline pupil size across blocks for different attentional states (top panel) and stickiness levels (bottom panel). The right plots show the estimated baseline pupil size from the best-fitting GAMM. Error bars in the right plots reflect the estimated 95% confidence intervals.

baseline pupil size was predictive of experiencing a specific attentional state. As shown in Figure 3.9, bottom panel and assessed with a GAMM, we also failed to find consistent differences in baseline pupil size between the different stickiness levels.

3.3.4 Task-evoked response in pupil size

We assessed the task-evoked response in pupil size for each attentional state and stickiness level separately for go and no-go trials. For all following analyses, we only considered correct trials. We present the grand averages of the task-evoked pupil

Figure 3.10. Task-evoked response in pupil size aligned to go stimulus onset ($t = 0$). The left plots show the average evoked response for each attentional state (top) and stickiness level (bottom) as observed in the data. The right plots show the estimates of the best-fitting GAMM models. We checked for significant differences between two evoked responses by determining a difference curve based on the estimated evoked responses. Two evoked responses were considered to be significantly different at a particular point in time when the pointwise 95% confidence interval around the estimated difference curve not include zero. We indicated a significant difference between two evoked responses with a colored bar.

responses along with the estimates of a fitted GAMM model in Figures 3.10 and 3.11, for go and no-go trials respectively.

3.3.4.1 Go trials

What is noticeable from the task-evoked pupil responses on go trials is that there appear to be two peaks in the pupil response. The first peak occurs at around 700 ms, followed

Figure 3.11. Task-evoked response in pupil size aligned to no-go stimulus onset ($t = 0$). The left plots show the average evoked response for each attentional state (top) and stickiness level (bottom) as observed in the data. The right plots show the estimates of the best-fitting GAMM models. Two evoked responses were considered to be significantly different at a particular point in time when the pointwise 95% confidence interval around the estimated difference curve not include zero. Significant differences between two curves were indicated with colored bars in the plot.

by a second peak at approximately 1200 ms. Although it is difficult to determine what is precisely reflected in these two peaks, it is reasonable to assume that the first peak reflects the amount of attention allocated to the (visual) processing of the stimulus, while the second peak may reflect processing related to the response and/or processing of the mask or fixation cross. Our results showed that the evoked response in pupil size was smaller at the first peak, but not at the second peak, when participants were engaged in self-generated thought ($t = [434 - 788 \text{ ms}]$) or other distractions ($t = [333 - 939 \text{ ms}]$)

compared to being on-task. For task-related interference we found no significant difference in the task-evoked pupil response from on-task.

With respect to the stickiness of thought, we found that the task-evoked response in pupil size was smaller when participants experienced sticky thought compared to neutral thought ($t = [283 - 1419 \text{ ms}]$), as well as non-sticky thought ($t = [611 - 737; 965 - 1167 \text{ ms}]$). However, the task-evoked response during non-sticky thought was not found to differ from the response during neutral thought.

3.3.4.2 No-go trials

Similar to the go trials, we found that the evoked response in pupil size on no-go trials was characterized by two peaks occurring at around the same time points as we observed for the go trials. Self-generated thought was found to be associated with a substantially smaller response in pupil size compared to on-task for the majority of the response ($t = [384 - 2000 \text{ ms}]$). For the other distracted states, we found no significant differences in pupil size compared to being on-task.

When participants reported having sticky thoughts, we found that the task-evoked response in pupil size was significantly smaller compared to neutral thought ($t = [510 - 864; 914 - 1672 \text{ ms}]$). Also for non-sticky thoughts we found that the evoked response in pupil size was smaller compared to neutral thought, but this difference only reached significance at the second peak ($t = [1192 - 1823 \text{ ms}]$). The difference in evoked response between sticky and non-sticky thoughts was not found to be significant at any timepoint.

3.4 DISCUSSION

The goal of this research was to explore the “stickiness” dimension of ongoing thought, which reflects a participants’ experienced difficulty of disengaging from thought (van Vugt & Broers, 2016; see also Joormann et al., 2011). We investigated how self-reported stickiness was associated with the participant’s attentional state, how it influenced task performance, and how it influenced pupil size. We adopted a variation of a sustained attention to response task (SART), which has been shown to be sensitive to lapses of attention (Cheyne, Solman, Carriere, & Smilek, 2009; Smallwood et al., 2004; Van Den Brink, Murphy, & Nieuwenhuis, 2016). Personal concerns of the participants were embedded in the SART to potentially increase the probability of observing sticky thought (Klinger, 2013; McVay & Kane, 2013).

3.4.1 Correlates and insights for the stickiness dimension of thought

We found that when participants reported having sticky thoughts, they also frequently reported being disengaged from the task (see also van Vugt & Broers, 2016). Conversely,

non-sticky thought (i.e., easy to disengage) and neutral thought (i.e., neither hard nor easy to disengage) were mostly associated with being focused on the task. Therefore, the results of the present experiment demonstrated that – at least in the context of sustained attention – ongoing thought is frequently experienced as difficult to disengage from off-task thought, but easy to withdraw from task focus.

On go trials, we found that reports of sticky thought could be discriminated from neutral or sticky thought in task-evoked pupil dilation, but not in behavioral indices. In contrast to earlier studies (see e.g., Bastian & Sackur, 2013; van Vugt & Broers, 2016), this research did not demonstrate faster response times (RT) and higher variance in response times (RTCV) to go trials when participants engaged in sticky, off-task, thinking. The absence of this effect was not an issue of power. Calculating Bayes factors separately for RT and RTCV demonstrated that the present study provides strong evidence for similar RT ($BF_{01} = 37.3$) and RTCV ($BF_{01} = 26.2$) across different degrees of sticky thinking. An explanation for the present results may be in the relationship between RTCV and the degree to which participants were disengaged from the task (see Cheyne et al., 2009). Increases in RTCV have been associated with a state of “tuning out”, where attention is partially allocated away the task while awareness to the general task context remains. The transient disengagement from the task during tuning out results in slowing and speeding of response times could lead to higher RTCV. In this experiment, participants were likely to be more strongly disengaged from the task during sticky thoughts – a state of “zoning out”. According to Cheyne et al. (2009), zoning out is associated with reactive and automatic responding to the task. It could be that the response time patterns as a result of automatic responding are not (measurably) different from responding during task focus (see also Jubera-García et al., 2019).

While behavioral indices were similar, task-evoked responses did differ. We observed a smaller task-evoked response in pupil size for go trials during episodes of sticky thought, suggesting that less attention is allocated to task processing compared to during episodes of neutral or non-sticky thought. Hence, sticky thought can be detected by looking for signs of perceptual decoupling in task-evoked pupil dilation, even at a time when behavior does not appear to suffer.

While behavioral indices could not distinguish between sticky and non-sticky thought in go trials, accuracy on no-go trials did discriminate between different levels of stickiness. Participants demonstrated a higher no-go accuracy (i.e., more often withheld a response) when they reported having non-sticky thought compared to neutral thought, but performed severely worse when they experienced sticky thought. In addition to the performance decrement with sticky thinking, we observed that task-evoked pupil

responses were smaller on correct no-go trials. Together with the smaller evoked response to go trials with sticky thought, this provides further evidence that sticky thought limits attention and exertion of cognitive control to external task processing (even when the response ends up being correct). In contrast, we could not discriminate non-sticky from neutral thought in task-evoked pupil dilation. Therefore, it is unclear whether cognitive processing leading to accurate performance differed when experiencing non-sticky or neutral thought, while average accuracy did differ. We argue that this indicates that participants could not reliably classify their thought as non-sticky or neutral. Instead, non-sticky reports may have been motivated by accuracy on the preceding no-go trial, explaining the better performance with non-sticky reports. Reports on sticky thought were likely not, or at least less, affected by no-go performance, since task-evoked pupil dilation was affected in correct trials.

The differences in no-go accuracy between sticky and neutral/non-sticky thought may provide some insight in how cognitive processing differs between these modes of thought. According to the literature on the SART, deliberate control is beneficial for performance on no-go trials. Deliberate control can be employed to sustain attention to the task (e.g., Manly, Robertson, Galloway, & Hawkins, 1999), but also to support a controlled response strategy (Dang, Figueroa, & Helton, 2018; Finkbeiner, Wilson, Russell, & Helton, 2015; Hiatt & Trafton, 2015). Therefore, the reason why sticky thought was associated with lower performance compared to neutral/non-sticky thought may be because this mode of thought was associated with a lower level of deliberate control.

How deliberate control influences the stickiness of thought, as well as other mechanisms outside of deliberate control, may be explained by the dynamic framework of spontaneous thought (see Christoff et al., 2016). This framework posits that the flow of content and orientation of thought can be constrained either through cognitive control (referred to as ‘deliberate constraints’), or through sensory and affective salience (referred to as ‘automatic constraints’). Deliberate constraints result in a neutral/non-sticky experience, because there is volitional control on the content and orientation of thought. On the other hand, strong automatic constraints (together with weak deliberate constraints) result in high stability of thought, which may make the thoughts difficult to disengage from.

The proposed relationship between the relative contribution of automatic and deliberate constraints to the stickiness of thought is supported by existing neuroimaging studies. Individuals with depression – a disorder marked by negatively valenced sticky thought (Nolen-Hoeksema, 2000; Nolen-Hoeksema et al., 2008) – have shown to have greater activation of the default network when engaged in experimental tasks compared to

controls (Anticevic et al., 2013; Whitfield-Gabrieli & Ford, 2012). The default network is proposed to support spontaneous thought (Christoff, 2012; Ellamil et al., 2016; Fox et al., 2015). The increased default network activity is, furthermore, accompanied with greater activation of (emotional) salience networks, while areas associated with cognitive control have reduced activation (Hamilton et al., 2012). Therefore, these results are consistent with the idea that sticky thought is mostly constrained through salience, and less so through deliberate control.

Chapter 4

Mechanisms of adaptive thought

In this chapter we reviewed several literatures to investigate the question how and why self-generated thought can be an adaptive process.

ABSTRACT

Many studies have indicated that engaging in self-generated thought can have an adaptive function. However, the question how and why that is has currently been largely neglected. Here, we aimed to answer this question by reviewing several, hitherto separate, literatures. We propose that self-generated thought can be adaptive, first of all, because our memory prioritizes the retrieval of memories with future relevance, predicted by our experiences in the past and current (unfinished) goals. By prioritizing the retrieval of memories, we may selectively strengthen memory structures that have likely future relevance, essentially preparing our memory for the future. In addition, we may also use the retrieved memories to build new memory structures in working memory. Building memory structures facilitates adaptive activities such as planning, but may also contribute to creative problem solving and decision making.

Keywords: self-generated thought; adaptive behavior; memory retrieval; rational memory; prospective memory; planning

4.1 INTRODUCTION

In our daily lives our thoughts are often shaped by the events in our environment. When we are driving and hit a crossing, we look around for current traffic and use this information to assess whether it is safe to cross or not. Nonetheless, our thinking is not merely bound to the here and now. On a long and straight highway, where driving imposes only low cognitive demands, we may start to entertain thoughts encompassing other places, events, and times. Also while reading this article, we may spontaneously engender thoughts that are unrelated to its content. Can engaging in such *self-generated thought* be rational and adaptive?

It is tempting to consider self-generated thought merely as a distraction and its content as useless. However, research is showing that self-generated thought is not useless at all. It has commonly been demonstrated that self-generated thought is frequently oriented to our personal futures, goals, and concerns (Baird et al., 2011; Kane et al., 2017; Klinger, Barta, & Maxeiner, 1980; Smallwood, Schooler, et al., 2011). Moreover, the content of such thought has found to be realistic, structured, and concrete (Stawarczyk et al., 2013). Many studies until now have indicated that self-generated thought can facilitate adaptive (goal-oriented) behavior. Engaging in self-generated thought has been shown to among others enhance creative insight and problem-solving (e.g., Baird et al., 2012; Gable, Hopper, & Schooler, 2019; Leszczynski et al., 2017; Tan, Zou, Chen, & Luo, 2015; Vanucci & Agnoli, 2019), to improve memory for personally relevant information (Smallwood, Schooler, et al., 2011; Steindorf & Rummel, 2017), to provide opportunities for future planning (Baird et al., 2011; Medea et al., 2016), to foster patience for long-term rewards in decision making (Smallwood, Ruby, & Singer, 2013), and to improve metacognitive monitoring (Allen et al., 2013). These works highlight at least two important things. First of all, the content of self-generated thought is not random. Instead, relevant and useful memories are seemingly *prioritized* for self-generated thought. Furthermore, it implies that engaging in self-generated thought is in fact useful and adaptive in some circumstances.

Although it is established that self-generated thought can facilitate adaptive and goal-directed behavior, we currently lack an account on how and why that is. A reason for that is that existing research on self-generated thought has been less concerned with the actual content of episodes, how memories are selected for self-generated thought, and what processes underlying self-generated thought are accountable for the observed adaptive outcomes (Berntsen, 2019). Therefore, we set out to review several, and hitherto mostly separate, literatures to address the question how and why self-generated thought can facilitate adaptive behavior. We argue that if we want to answer the question how and why self-generated thought can have an adaptive

function, a logical first step is to consider what makes us think about relevant and useful memories in the first place. The experience of self-generated thought critically depends on the retrieval of experiences from memory (see e.g., Andrews-Hanna et al., 2014; Fox, Nathan Spreng, Ellamil, Andrews-Hanna, & Christoff, 2015; Mills, Herrera-Bennett, Faber, & Christoff, 2018). Therefore, we look into the memory literature to discuss what mechanisms and factors influence the selection of memories for self-generated thought and *prioritize* its content to relevant and useful memories. The second step is to examine how the prioritized selection of memories, in addition to the processes that determine what we do with the selected memories, can facilitate adaptive behavior. To limit the scope of the review, we decided to focus on planning, since planning is most likely to be the most prominent adaptive activity during self-generated thought (Baird et al., 2011). However, as we will discuss later in the conclusion section, the proposed mechanisms that support planning may also be applicable to other adaptive activities such as creative problem solving and decision making.

4.2 SELECTION AND PRIORITIZATION OF EXPERIENCES FOR SELF-GENERATED THOUGHT

4.2.1 Rational selection of experiences for self-generated thought

Early insights on the prioritization of memory retrieval are provided by research on the rationality of human declarative memory. Anderson and Schooler proposed that human memory is *rational*. With that, the researchers meant that our memory always tries to provide the most relevant and useful window to our past by remembering what we likely need again, and by forgetting what has outlived its usefulness. They observed that the likelihood that memories are retrieved, also known as the *accessibility of memories*, was associated with the probability that such memories were currently needed (Anderson & Schooler, 1991; Schooler & Anderson, 1997). Such odds were learned from statistical patterns in past experience, notably the *recency* of exposure, the *frequency* of exposure, and the *context* of prior exposures. To illustrate, memories of events that just occurred or happened frequently are likely still relevant now. We can easily remember the name of a good friend or of a person we have just met (see e.g., Baddeley & Hitch, 1993; Murdock, 1962). In addition, context is an important determinant of what is currently relevant. We are more likely to recall a grocery shopping list if we are in a supermarket than in an unrelated context (see e.g., Godden & Baddeley, 1975; Macken, 2002; Smith & Vela, 2001).

The rational memory theory puts forward a factor that prioritizes memories for self-generated thought: the estimated *need* of a memory, given the current circumstances and history of past relevance and usefulness. If self-generated thought indeed depends on rational memory retrieval processes, what does this predict about its content? First of all, self-generated thought should preferentially include thoughts about recently

and/or frequently encountered experiences. This prediction is supported by research demonstrating that self-generated thought mostly encompasses recent past and near future (Spronken, Holland, Figner, & Dijksterhuis, 2016; Stawarczyk et al., 2011). Furthermore, it is consistent with studies showing that frequently manifesting concerns are important determinants of self-generated thought (Stawarczyk et al., 2013). One would also expect that environmental information (i.e., external context) and current working memory content (i.e., internal context) would bias the content of self-generated thought to associated information. Consistent with this prediction, Kvavilashvili and Fisher (2007) showed that spontaneous thoughts about delayed intentions were often reported following subtle cues associated with the intention. Also, recent work showed that performing self-referential judgements (“Am I kind?”) instigated spontaneous mental elaborations and mind wandering (Daamen, van Vugt, & Taatgen, 2016; Huijser et al., 2018). Together, this indicates that there might be a rational basis for the content of self-generated thought. In other words, that statistical patterns in our past experiences influence what memories will most likely be retrieved during self-generated thought.

4.2.2 Suspended goals and concerns prioritize the selection of future relevant memories

As discussed so far, the rational memory theory explains that the current need of memory is predicted by its past. However, how can we reconcile this with the observation that self-generated thought is frequently oriented to the future? Put differently, how can memories with *future* relevance also be strongly accessible in the *present*?

The answer to the question above is that currently suspended goals prioritize the selection of its associated memories. Goals are represented in memory, and can therefore spread activation to memories that are associated with it (Altmann & Trafton, 2002; Goschke & Kuhl, 1993). As a result, goals can influence the need of associated memories, and therefore, bias the content of self-generated thought.

Support for the influence of suspended goals on associated memories comes from research showing that material that is associated with a goal is retrieved more quickly and reliably than material with no related goal (e.g., Goschke & Kuhl, 1993; Lebiere & Lee, 2002; Marsh, Hicks, & Bryan, 1999; Meilán, Carro, Arana, & Pérez, 2011; Schult & Steffens, 2017). In the context of self-generated thought, Steindorf and Rummel (2017) demonstrated that suspended goals influence the content of our self-generated thoughts. The researchers performed an experiment in which the participants studied a grocery shopping list of 20 items for cued recall. After recall of half of the list, one group of participants were told that they had to recall the other half of the list after an interleaving task (‘interrupted’ condition). The other group was told that the recall

task was finished ('finished' condition). Note that the rational memory theory would not predict a difference in the content of self-generated thought between the groups. The study phase was the same for both, therefore the items have similar recency and frequency in memory. However, the results of the study showed that participants in the interrupted condition consistently used self-generated thought as an opportunity to rehearse the items. This was not found in the finished condition, although the overall frequency of self-generated thought (irrespective of the content) was the same. This study, therefore, directly demonstrated that suspended goals can prioritize which memories are selected for self-generated thought. Furthermore, it demonstrated that as soon as a goal is finished, we quickly stop thinking about the goal (see also Scullin, Bugg, McDaniel, & Einstein, 2011).

4.2.2.1 *Prioritization between suspended goals*

Goals help to select memories with future relevance for self-generated thought. However, how do we prioritize when we have multiple goals? There is some evidence that—not surprisingly—those goals we subjectively consider to be more important affect spontaneous thought more strongly. For example, Szarras and Niedźwieńska (2011) demonstrated that participants had more deliberate and spontaneous thoughts about planned tasks that were reported as more important. Klinger, Barta, and Maxeiner (1980) showed that the frequency of thought about a particular goal can best be predicted by its anticipated value, probability of success, and urgency. Therefore, this suggests that we may prioritize our goals based on their desirability, feasibility, and imminence.

4.2.2.2 *Goals can be cued by context*

Apart from being cued by subjective importance, goals can also be cued by the current context. Research on prospective memory (i.e., memory for postponed intentions) has demonstrated that a substantial proportion of future- and goal-oriented thought (40-70%) is triggered accidentally by internal cues ('what happens to be in one's thought') and environmental cues (see Kvavilashvili & Fisher, 2007; Szarras & Niedźwieńska, 2011). These observations are explained by the fact that processing cues can enhance the memory accessibility of an associated action or actions (McDaniel et al., 2015; see also Huijser, van Vugt, & Taatgen, 2018). To illustrate, encountering a phone in our office makes it more likely to (spontaneously) retrieve memory of the intended action to call a colleague. When the current circumstances allow for the execution of the intended action, we respond accordingly. If not, we may respond with self-generated thought (see Klinger & Miles Cox, 2011).

The prospective memory literature predicts that future-oriented self-generated thought

is often triggered by context, because cues enhance the memory accessibility of associated intentions. If we apply the rational memory theory to this, we may also understand this as that we engage in future-oriented self-generated thought when the current context predicts the need of an intention. Interestingly, this implies that how intentions are encoded in memory may affect the likelihood that we engage in specific self-generated thought. Tentative support for this hypothesis is the observation that intentions encoded as *if (context)/ then (action)* – plans are more likely to be successfully remembered (McCrea et al., 2017; Rummel, Einstein, & Rampey, 2012). When we think about the future, this in itself may also increase the likelihood for future- and goal-oriented self-generated thought. Szarras and Niedźwieńska (2011) observed that intentions that were initially recalled resulted in higher frequency of spontaneous rehearsal without any triggers and deliberate rehearsals later in the delay period. In light of the rational memory theory, this can be explained by the increased frequency and renewed recency of the recalled intention in memory. In essence, self-generated thought may therefore play a role in preparing our memory for the future by enhancing the accessibility of future relevant memories.

4.3 MECHANISMS UNDERLYING THE ADAPTIVENESS OF SELF-GENERATED THOUGHT

Now that we have discussed how memories may be selected and prioritized for self-generated thought, we can focus on the possible mechanisms that are responsible for the adaptiveness of self-generated thinking. What is it that we do with the retrieved memories that makes self-generated thought frequently adaptive?

Arguably, a core aspect of self-generated thought is the ability to simulate experiences. We do not just rehearse separate memories. Instead, it is known from the episodic future thinking literature that we sample multiple memories from our relevant and useful past, and flexibly (re-)combine those memories to form novel experiences (i.e., constructive simulation hypothesis; see e.g., Addis & Schacter, 2008; Hassabis & Maguire, 2007; Schacter & Addis, 2009; Schacter, Addis, & Szpunar, 2017; Schacter, Norman, & Koutstaal, 1998). The ability to generate novel mental simulations is helpful, first of all, because it allows us to strengthen (potentially) future relevant memory structures (Antony, Ferreira, Norman, & Wimber, 2017; Steindorf & Rummel, 2017). Thereby, we increase the likelihood of remembering relevant memory structures as soon as they are needed. Apart from strengthening existing memory structures, a second important mechanism is that we may draw novel connections between the retrieved memories, resulting in updated or new memory structures. Learning the novel connections is suggested to depend on working memory (see Oberauer, 2005; Oberauer, Souza, Druey, & Gade, 2013). All in all, the adaptiveness of self-generated thought may therefore be in strengthening and constructing of memory structures that have potential usefulness for our future (see

also Christoff, Gordon, & Smith, 2011).

4.3.1 Planning: building memory structures for the future

Of all the adaptive activities that we may perform during self-generated thought, planning is an activity in which building memory structures plays an integral role (e.g., Cole, Bagic, Kass, & Schneider, 2010; Cole, Patrick, Meiran, & Braver, 2018; Oberauer, Souza, Druet, & Gade, 2013; Pereg & Meiran, 2019). It is also an activity that is prominently performed during self-generated thought, in particular ‘autobiographical’ planning (see Baird et al., 2011b). Autobiographical planning is defined as a process in which we identify the necessary steps towards achieving personally relevant future goals (Spreng, Gerlach, Turner, & Schacter, 2015). Examples of personal goals for which we can plan include to prepare a vacation, to develop a story for a talk, and to finish a PhD project. Depending on the goal, planning would involve identifying and sequencing which tasks, operations, and/or factual concepts are needed. Note that the elements of the plan are retrieved from long-term memory (see also Mayr & Kliegl, 2000). We bind the elements in working memory, to form a hierarchical representation of the planned activity, which we may later implement to guide towards goal achievement.

The requirement for working memory resources in the planning process has several implications. First of all, we may expect that we can only plan during self-generated thought (without interference) during moments when working memory is not already occupied (see Nijboer, Borst, van Rijn, & Taatgen, 2016; Salvucci & Taatgen, 2008; Wickens, 2002). Recently, we investigated how performing a present task influences to what extent people are able to plan ahead (i.e., build memory structures for future tasks) for an upcoming task (Huijser, Taatgen, & van Vugt, 2021; Chapter 5 of this thesis). Participants performed a sequential dual-task paradigm in which they could (voluntarily) prepare for a second task while performing a first task. We manipulated working memory load in the first task on various levels (i.e., holding zero, one, or two items in memory), informed by the earlier mentioned work indicating that planning depends on working memory to build the task representation. The results showed that participants could productively plan ahead with little cost to the first task. Perhaps surprisingly, holding one item in memory for the first task did not hamper planning much. However, maintaining a list of two items did. Holding a list of two items in memory requires representing a novel binding between two stimuli. We argue that the interference for holding two items to planning, in contrast to needing to maintain a single item, provides support for the idea that working memory is involved in representing the novel bindings in memory structures.

Another implication of the role of working memory in building the memory structures

is that working memory bindings are strong, but also very fragile. In other words, we can quickly learn new associations between memories to form new structures, but these associations may also be quickly forgotten. This raises questions about the actual effectiveness of adaptive self-generated thought, in particular when there is a long delay between planning for a goal and its execution. The prospective memory literature has identified two pathways towards successful prospective remembering (see e.g., McDaniel et al., 2015). The first pathway to successful prospective remembering is to actively maintain the plan in memory (see also Altmann & Trafton, 2002). As we already discussed in the previous section, self-generated thought may play a role in this through facilitating rehearsal of the to-be-maintained plan. The second pathway relies on spontaneous retrieval. When cues have been associated to memory structures, we may spontaneously retrieve the memory structures upon encountering cues associated with it in our environment.

4.4 CONCLUSIONS AND OPEN QUESTIONS

Self-generated thought has long been acknowledged to be predominantly oriented to our future and to have an adaptive function. Nevertheless, we currently do not have an account of how and why that is. Based on the reviewed literature, we proposed that self-generated thought can have an adaptive function because our memory prioritizes the selection of (currently) relevant and useful memories, where relevance and utility is predicted by our past and current goals. With the retrieved experiences we may subsequently strengthen, update, and construct memory structures to improve our memory for postponed intentions and to construct plans for future tasks.

4.4.1 Beyond planning: creative problem solving and decision making

So far, we have primarily focused on planning to illustrate on how building memory structures leads to adaptive self-generated thought. However, the same mechanism of establishing connections in working memory may also support other adaptive activities such as creative problem solving and decision making.

A role for self-generated thinking in creative problem solving has long been suggested (see e.g., Baird et al., 2012; Campbell, 1960; Christoff et al., 2011; de Dreu, Nijstad, Baas, Wolsink, & Roskes, 2012; Fox & Beaty, 2019). Self-generated thought may support creativity, because its content is frequently generated in the absence of strong constraints, allowing for thoughts that are not or only weakly associated (see Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016). For such thinking to be useful for problem solving, however, it is necessary to draw connections between these loosely associated memories and to temporarily maintain them to build a memory structure representing the solution. Similar to planning, these novel connections are likely constructed in

working memory.

In contrast to planning and creative problem solving, the potential contribution of self-generated thought to decision making has currently received little attention. To our knowledge, only Smallwood and colleagues (2013) examined for a role of self-generated thought in decision making. The researchers demonstrated that engaging in self-generated thought was associated with being more patient for distant rewards (i.e., smaller delay discounting). However, we argue that the role of self-generated thought in decision making may be broader. Specifically, self-generated thought may facilitate decision making between alternatives that we have not directly experienced before, or involve uncertainty. In such cases, research has indicated that we need to integrate memories of multiple past experiences to estimate the value of the options (Daw & Dayan, 2014; Daw, Niv, & Dayan, 2005; Shohamy & Daw, 2015). It is often assumed that we perform such integration of memories at decision time. However, current research indicates that we also integrate memories offline, that is, prior to the decision is faced (Schuck & Niv, 2019). For example, Momennejad and colleagues (2018) found that in a multi-step decision task, memories of decisions in early steps are reactivated (determined by pattern classification of fMRI data) in offline periods following decisions later in the trajectory (see also Dohmatob, Dumas, & Bzdok, 2020; Gershman, Markman, & Otto, 2014). The frequency of such reactivations was predictive of how much participants later adapted their choice on the early decisions, indicating transfer of knowledge from decision steps later in the trajectory to early steps. The role of self-generated thought in such offline memory integration has to be still confirmed. Adding thought probing to such experimental paradigms could be an interesting starting point for future investigations.

4.4.2 Open questions

We have discussed how and why self-generated thought can be adaptive. However, something we have not discussed is how individual differences in cognitive function may influence the adaptiveness of self-generated thinking. Trying to plan ahead when the present task demands working memory resources may result in interference, costs to the present task, and reduced benefits from planning (see e.g., Huijser et al., 2021; Levinson, Smallwood, & Davidson, 2012b; Pereg & Meiran, 2019; Smallwood, Brown, et al., 2011). Research has indicated that people differ in how well they regulate their self-generated thought, and adjust it based on the current task context. Individuals with high working memory capacity have been shown to be less prone to engaging in self-generated thought when the present task is demanding compared to individuals who have relative low working memory capacity (see Rummel & Boywitt, 2014). Consequently, individuals with good working memory capacity may be good at picking the right moments to engage in adaptive self-generated thought, and therefore, benefit from it. However,

individuals with low working memory capacity may tend to engage in it more when it is inappropriate, resulting in more costs, and reduced benefit. We encourage future research to investigate the relationship between individual differences in the regulation of self-generated thought and its adaptiveness.

In addition, something we only slightly touched upon is how the selection of experiences for self-generated thought may be prioritized on the basis of their associated value. We mentioned that goals enhance the accessibility of its related memories, and that goals are prioritized based in part on their expected value. However, there are currently also indications that we prioritize memories by value directly, irrespective of goals. Bear and colleagues (2020) investigated what (spontaneously) comes to mind when you give people statements such as ‘how many hours of sports a week’. With a series of experiments the researchers demonstrated that what people sample from memory is a compromise of what is both likely (based on frequency statistics) and good (based on prescriptive value). In other words, they found that participants selected an answer based on the product of need and value. It should be noted that it is currently unclear what kind of value is at work here (e.g., prescriptive value, expected value, surprise, abstract reward of related to goal closure etc.). Furthermore, it remains to be investigated whether experiences for self-generated thought are also selected based on need and value.

4.4.2 Conclusion

In short, by combining separate lines of research we found that the adaptiveness of self-generated thought may – first of all – be explained by the rational prioritization of memory retrieval through our past, and the influence of unfinished goals. With the retrieval of such memories, we can prepare our memory for the future. Furthermore, we may build new memory structures in working memory, supporting planning activities, and potentially creative problem solving and decision making.

Chapter 5

The art of planning ahead

In this chapter we investigated how demand from an ongoing activity can influence to what extent we are able to productively plan ahead for an upcoming task.

This chapter has been previously published as:

Huijser, S., Taatgen, N. A., & van Vugt, M. K. (2021). The art of planning ahead: When do we prepare for the future and when is it effective? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(5), 705-726. <https://doi.org/10.1037/xlm0000970>

ABSTRACT

Preparing for the future during ongoing activities is an essential skill. Yet, it is currently unclear to what extent we can prepare for the future in parallel with another task. In two experiments, we investigated how characteristics of a present task influenced whether and when participants prepared for the future, as well as its usefulness. We focused on the influence of concurrent working memory load, assuming that working memory would interfere most strongly with preparation. In both experiments, participants performed a novel sequential dual-task paradigm, in which they could voluntarily prepare for a second task while performing a first task. We identified task preparation by means of eye tracking, by detecting when participants switched their gaze to information about the second task while performing the first task. The results showed that participants prepared, but also that there were large individual differences in how often they did so. When participants prepared, it was productive, as evidenced by faster RTs on the second task and only a small cost to the present task. The probability of preparation and its productiveness decreased with increases in the difficulty of the first task. In particular, we found that working memory load from the first task interfered with preparation. We conclude from our study that people can productively prepare for the future whilst performing an ongoing task, and that it is possible to track this preparation process empirically. In addition, we conclude that working memory resources play an important role in task preparation.

Keywords: Task preparation, planning, working memory, rapid instructed task learning, eye tracking.

5.1 INTRODUCTION

Task preparation is a complex activity in which we build a task representation from acquired skills, concepts, and facts to achieve a future task goal (Morris & Ward, 2005; see also Cole, Laurent, & Stocco, 2013). Although complex, preparing for upcoming tasks is arguably an essential activity in our daily functioning. Many tasks that we perform require preparation. Whether it is to cook an evening dinner, navigating to a novel location, or writing a paper like this one, we always need to think about how we are going to do it.

Preparing in advance often means that we have to perform it alongside an ongoing activity. For example, we might plan the structure of a paper while commuting to work. Whether such preparations are useful may depend on the nature of the current ongoing task. Planning is likely to be efficient and effective when demands from traffic during the commute are low. However, when traffic becomes demanding, such as at busy crossing, preparing for the future may become hard and potentially useless. Therefore, when we need to plan concurrently with a present task, it is important that we can decide correctly when it is a suitable moment to prepare or not. Interesting questions are which factors determine suitable moments for preparation in relation to the current task, but also to what extent preparation pays off, and whether preparing for the future hurts the present task. Research studying these questions is currently limited.

5.1.1 Task preparation in the mind wandering and prospective memory literature

Insights into planning and specifically task preparation come largely from two related fields of research, namely: mind wandering and prospective memory. Mind wandering is of particular interest here, because it refers to thought processes that occur in the context of ongoing activities, but are unrelated to these activities (Smallwood & Schooler, 2015). Many studies have demonstrated that mind wandering is not merely a distraction, but that a large proportion of our mind wandering involves planning for future tasks (e.g., Baird, Smallwood, & Schooler, 2011; Kane et al., 2017; Smallwood, Nind, & Connor, 2009; Stawarczyk et al., 2013; Stawarczyk, Majerus, Maj, Van der Linden, & D'Argembeau, 2011; van Vugt & Broers, 2016). Whether we engage in mind wandering depends on the difficulty of the present task (e.g., Feng, D'Mello, & Graesser, 2013; Seli, Risko, & Smilek, 2016). In particular, present tasks that require working memory resources have been shown to decrease mind wandering frequency (e.g., Levinson et al., 2012; Smallwood et al., 2011, 2009). Researchers have argued that this is because the maintenance of mind wandering depends on access to working memory (Huijser et al., 2018; Smallwood & Schooler, 2006). Smallwood and colleagues (2011, 2009) showed that in particular future-oriented mind wandering, and therefore maybe planning, was

suppressed by working memory load compared to mind wandering about the past or the here and now. Mind wandering research thus suggests that the decision to engage in planning or task preparation may depend on whether the present task is using working memory.

Prospective memory research investigates whether planning ahead is effective by studying how we can remember to perform intended actions across periods of interfering activities (e.g., to hang out the laundry when coming home from work). The effectiveness of planning ahead depends on both strategic monitoring for the occurrence of relevant cues in the environment as well as spontaneous retrieval and rehearsal processes (Kvavilashvili & Fisher, 2007; McDaniel et al., 2015). The role of strategic monitoring in prospective memory has been supported by research showing that prospective memory performance is sensitive to the amount of available executive control resources. For example, Marsh, Hancock, and Hicks (2002) conducted a series of experiments to investigate how demand on executive control influences the likelihood of remembering to perform an intended action. The researchers manipulated demand on executive control resources by requiring one group of participants to coordinate two task rules for the ongoing task (i.e., switching group), whereas the other group of participants only applied a single task rule (i.e., control group). Prospective memory performance was assessed by embedding prospective memory cues in the ongoing task. Participants were instructed to press a specific key upon encountering a cue, indicating that they remembered the intended action. The results demonstrated that the switching group had significantly lower prospective memory performance compared to control group that did not need to swap rules for the ongoing task. In a follow-up experiment (see Experiment 3), the researchers showed that when the two task rules can be combined (i.e., removing the need to swap task rules), prospective memory performance was not different from the control group with a single task rule. Together, these results indicate that the necessity to coordinate multiple tasks can limit our capacity to remember and execute planned actions.

How we plan also influences the effectiveness of planning ahead. Prospective remembering is enhanced when a plan includes an explicit link between the future situation and the intended action(s) (i.e., “When I encounter X, I will do Y”) compared to when only the intended action is considered (e.g., “I need to do Y soon!”; Gollwitzer, 1996; McCrea et al., 2017; McDaniel, Howard, & Butler, 2008; Rummel, Einstein, & Rampey, 2012). Recent research demonstrated that we may generate such plans while being engaged in a present task (see Scullin et al., 2018). The researchers found that participants generated specific exemplars of when to execute the intended action during the task (e.g., to press ‘Q’ at encountering ‘apple’ when the cue is ‘fruits’) and

that generating such exemplars was beneficial to later prospective remembering. All in all, prospective memory research thus shows that planning ahead is effective, even when other activities come in between. Furthermore, planning processes may persist throughout ongoing activities and may be beneficial to later goal attainment.

Despite these insights from mind wandering and prospective memory research, those research fields do not provide a complete picture of *when* we engage in task preparation and *how* useful/costly such preparations are. Mind wandering research has mainly focused on *detecting* the occurrence of thoughts about planning during present tasks. Therefore, not much is known about what planning-related thought content during mind wandering entails, how those thought reflect the construction of actual plans, and how useful this planning is (see also Berntsen, 2019). Prospective memory research has mostly studied very simple intentions (e.g., pressing a different key whenever a target stimulus is presented). Therefore, it is unclear whether insights from this literature also apply to preparing for more complex tasks such as designing the structure of an article.

With this article, we want to make the first steps towards studying when and how task preparation takes place in a present task. We are interested in how the characteristics of a current task influence *whether* and *when* we engage in preparation. Furthermore, we want to understand how the current task influences to what extent preparation pays off for the future task, and whether it hurts the current task. To help us predict when preparation can be performed effectively in parallel with an ongoing task, i.e. when task preparation and ongoing task processes do not interfere with each other, it is helpful to review research on multitasking.

5.1.2 Multitasking: when can future task preparation co-occur with a present task?

In the multitasking literature, several ideas have been put forward to explain when two or more tasks interfere. First of all, it has been proposed that two tasks can in principle be performed at the same time, but interference may occur due to problems in scheduling the two tasks (e.g., Cooper & Shallice, 2000; Meyer & Kieras, 1997). Control mechanisms may determine how the tasks are interleaved or prioritize processing for one task, resulting in delayed processing for one of the tasks. When we apply this idea to concurrent task preparation, this means that how and when we can prepare for the future during a present task mainly depends on our control strategy.

In contrast to the previous claim that two tasks can in principle be concurrently performed, other researchers have claimed that human cognition is limited in capacity. Because of that, bottlenecks may occur when two tasks require similar cognitive

resources. Several (single) bottlenecks have been proposed, such as in perception (e.g., Broadbent, 1958), response selection (e.g., Pashler, 1984, 1994), and motor control (e.g., Keele, 1973). Multiple-resource theories of multitasking have tried to unify these different bottleneck accounts into a comprehensive framework (see e.g., Salvucci & Taatgen, 2008; Wickens, 2002). Threaded cognition is an example of a relatively recent multiple-resource theory (see e.g., Borst, Taatgen, & van Rijn, 2010; Nijboer, Borst, van Rijn, & Taatgen, 2016; Salvucci & Taatgen, 2010; Salvucci, Taatgen, & Borst, 2009). The threaded cognition theory argues that all resources in human cognition (e.g., vision, motor, procedural, memory) can act as bottlenecks during multitasking (see Salvucci & Taatgen, 2010). In addition, it claims that all resources can only be of service to a single task at a time. Threaded cognition therefore predicts that we can only prepare for an upcoming task during a present task when this task does not rely on the same resources. In other words, suitable moments for preparation occur when its required resources are not currently occupied by a main task. When resources need to be shared, preparation efforts may need to wait before the required resources are released, making the efforts inefficient and potentially ineffective.

Task preparation requires at least two types of cognitive resources: long-term memory, and working memory. Long-term memory is used to collect the necessary information for the construction of the task representation, such as facts, concepts, and required skills. The actual construction of the task representation happens in working memory. Working memory is used to represent the intermediate steps in the planning process, and helps to connect these intermediate steps to the rest of the action plan in long-term memory (Cole, Braver, & Meiran, 2017; see also Oberauer & Hein, 2012 for a similar interpretation of working memory). The likelihood of engaging in task preparation during a present task is therefore high when long-term memory and working memory resources are free. However, the likelihood of preparation is lower when one or more of these resources are occupied. Moreover, even if preparation occurs, its effectiveness will be lower when the required resources are not free.

5.1.3 The current study

In this study, we conduct two experiments to study how differential demands from a present task influence when we prepare and how efficient and effective our preparation efforts are. Demands will be manipulated by varying working memory load. Since working memory is required to construct task representations, we expect that working memory demands may be a critical factor in deciding when to prepare for the future or not. In addition, it also may determine the extent to which preparation is useful/costly.

In both experiments, we use a novel sequential dual-task paradigm. The core aspect of

this paradigm is that participants can voluntarily prepare for a second task while they are performing a first task. The first task involves responding to occasional probe digits in streams of digits. We manipulate demand on working memory by requiring participants to continuously maintain one or more digits during the stream. We refer to the first task in this paper as the ‘digit parity task.’ The second task that we use in our experiments is the rapid instructed task learning (RITL) paradigm (see e.g., Cole, Braver, & Meiran, 2017), which is a task perfectly suited for studying task preparation and planning. The reason for this is that every trial in the RITL task has different task instructions. A participant in this task has to read the instructions, formulate a task representation based on these instructions (i.e., prepare the upcoming task), and then apply this task representation to a set of stimuli. For example, upon reading the instruction “same, sweet, left-index”, participants need to interpret this as: “if the answer to ‘is it sweet’ is the same for both stimuli, then I press with my left-index finger.” If the stimuli are ‘Apple’ and ‘Pear’, the correct answer is pressing with the left index finger since both apple and pear are sweet.

There are several reasons why we think the RITL paradigm is suitable for our research goals. First of all, preparation in the RITL task is not open-ended. Previous research has shown that participants need around four seconds of preparation after seeing the task instructions (Cole et al., 2018). Secondly, it has been shown that performance in this task is sensitive to the amount of time spent on preparation. Lastly, the fact that it is possible to have different instructions for each trial means that we can collect repeated measures of task preparation. All in all, this suggests that the RITL task is well-suited for investigating whether preparation efforts are useful or not.

We want to measure the occurrence of preparation efforts while participants engage in the digit parity task. Measuring preparation is challenging because it is largely an internal and covert process. Other internal thought processes such as mind wandering have been commonly investigated by periodically asking participants to report on their current thought content (see Smallwood & Schooler, 2015; Weinstein, 2017). However, such methods lack the temporal precision to indicate when participants started to prepare. Other studies have also used self-caught methods, in which participants indicate themselves when they notice they are engaging in a certain internal thought process (e.g., Schooler, 2002; Seli et al., 2017). However, such self-caught methods require participants to constantly monitor their own thought, essentially introducing another task. Moreover, participants may not be very accurate in catching every episode of internal thought.

To overcome the issues of self-report methods, we use eye-tracking to measure future task preparation in both experiments. During the full duration of each trial, we

presented the RITL instructions in the periphery of the screen and measured preparation attempts by detecting eye-fixations to the instructions during the digit parity task. Since the instructions were unique for each trial, looking at the instructions indicates that participants are preparing for the RITL. Unlike self-report methods, glances at the instructions give us a reasonably precise indication of when participants start with preparing.

5.2 EXPERIMENT 1

We designed Experiment 1 to investigate how different levels of working memory load would affect when participants engage in voluntary preparation, and the usefulness of this preparation. To manipulate working memory load we created three conditions of the first digit parity task. The first condition involved no working memory load because participants did not need to maintain any of the digits in the stream. We called this condition the *no-load* condition. The second *low-load* condition required participants to maintain the last digit in the stream. Therefore, the low-load condition had a working memory load of one item. The third *high-load* condition asked participants to maintain the last two digits, therefore resulting in a working memory load of two items. We expected that the working memory load in the low- and high-load condition would act as a bottleneck for preparing for the upcoming rapid instructed task learning (RITL) task. Threaded cognition predicts that working memory cannot be shared between two activities, therefore, even maintaining a single item in working memory should already block preparation. Consequently, we expected to see less preparation in the low- and high-load condition compared to the no-load condition across several measures, including the probability of preparing in a trial, the amount of observed preparation attempts in a trial, and the duration of the preparation. Furthermore, we expected that preparation during low- and high-load trials would result in a smaller benefit to performance (i.e., error rate and response time) in the RITL task than preparation during the no-load trials.

5.2.1 Method

5.2.1.1 Participants

For Experiment 1, we recruited 38 participants (14 female; $M_{\text{age}} = 21.6$; $\text{range}_{\text{age}} = 18 - 34$) from the University of Groningen and the Hanze University of Applied Sciences. We screened the participants for having normal or corrected-to-normal vision prior to taking part in the experiment. We excluded participants wearing glasses or hard contact lenses since such corrective measures often result in tracking issues. All participants except one were native speakers of Dutch. Since being a native speaker of Dutch was a requirement for participation, we excluded the non-native speaker from all further analyses. All participants provided informed consent before the experiment and received a monetary compensation after testing (12.00 Euros for the 1.5-hour experiment). The experiment

Figure 5.1. (left) Example of the screen layout while performing the digit parity task. (right) Example of the screen layout while performing the rapid instructed task learning (RITL) task. The arrow indicates the transition of the screen layout when the experiment moved from the first to the second task. Note that the instruction for the RITL task was on the screen during both tasks. Text in the images is scaled larger for graphical purposes and translated from Dutch (original) to English.

was conducted in accordance with the Declaration of Helsinki and was approved by the research ethics committee of the Faculty of Arts, University of Groningen (CETO; research code: 61108926).

5.2.1.2 Experimental paradigm

This experiment used a sequential dual-task paradigm. Although the two tasks were performed sequentially, participants could already voluntarily prepare for the second task while they were performing the first task. The first task in the paradigm was a digit parity task. When finished, this task was followed by a rapid instructed task learning (RITL) task (adapted from Cole, Patrick, Meiran, & Braver, 2017). Each block in the experiment consisted of a single working memory load condition (no-load, low-load or high-load). The manipulations of load are described in the *Digit parity task* section below.

Each of the two tasks was presented within a separate window on the screen (650 x 650 pixels; 100 px vertical separation). We presented the digit parity task in the left window and the RITL task in the right window. We highlighted the window border in green to indicate which task was active during a trial. An example of the layout of the tasks can be found in Figure 5.1.

5.2.1.2.1 Digit parity task

In the digit parity task, participants were instructed to judge whether digits were odd or even. A sequence of ten randomly selected digits between 1 and 8 was presented on each trial (1500 ms), separated by inter-stimuli intervals (ISI) indicated with a “+” (1500

Figure 5.2. Overview of a single trial in this experiment. One trial consisted of two tasks: a digit parity task (upper part in the figure) and a rapid instructed task learning (RITL) task (bottom part). The digit parity task had three versions: one placing no load on working memory (o-back), one placing low load on working memory (1-back), and one placing high load on working memory (2-back). Only one version was performed in each block of trials. Text in the experiment was in Dutch, but translated in this figure to English to facilitate understanding.

ms). A random sample of the ten digits was highlighted with a symbol. We called these digits: *probes*. The other digit stimuli were *non-probes*. Participants were instructed to perform a parity judgement whenever a probe was presented. No action was required for non-probes.

Depending on the working memory load version of the task, the parity judgment was performed on the currently presented digit and probe (*no-load*), on the digit prior to the probe (*low-load*), or on the digit two steps back in the sequence (*high-load*). We used a different symbol for the probe (i.e., a circle, square, or triangle) to indicate the version of the task. The symbols were counterbalanced by participant number. In all versions, participants indicated their answer by pressing the ‘m’ key on the keyboard for *even* and ‘n’ for *odd*. We allowed responses during the presentation of the probe, therefore limiting responses to 1500 ms. After response, the probe was removed from the screen and replaced by the inter-stimulus interval. We added the remaining time of the probe (1500 – response time) to the duration of the following interval to make sure that the combined duration of the stimulus and ISI was kept at 3000 ms. Feedback was provided after each response or missed response with a different sound indicating correct or incorrect/missing responses (both approx. 65 dB). Volume was kept constant across participants.

The low- and high-load version of the task always required a history of one or two digits respectively. Therefore, we provided one or two digits (always non-probes) in advance of the trial on low- and high-load trials. Unlike the main digit sequence, these digits were presented at the center of a blank black screen. In the case of high-load trials, the two digits were presented simultaneously, separated by a “+”. Participants could start with the trial by pressing space. In no-load trials, this screen only asked participants to press space to start. After pressing space, the two windows for the two tasks were drawn on the screen, with the instruction for the RITL task in the right window, and the left window blank and highlighted (see Figure 5.1). After 200 ms, the first digit stimulus was presented in the left window. See Figure 5.2 for a graphical overview of a single trial for all versions of the digit parity task.

5.2.1.2.2 *Rapid instructed task learning task*

As soon as digit parity task was finished, the border of the left window turned white and the right turned green (see right panel in Figure 5.1). Alongside the change in color, a plus symbol “+” equal to the ISI in the digit parity task was drawn at the center of the right window. The instructions for the RITL task were already visible in the right upper corner of the window since the start of the digit parity task. We wanted to make sure that the first stimuli for the RITL task were presented only when people had switched tasks. This

was important to guarantee valid and accurate response times on these first stimuli. To achieve this, we presented the first stimuli as soon as the participants' gaze was detected around the plus symbol (200x150 pixel window) or instruction (200x150 pixels) in the upper right corner. To prevent the experiment from stalling and participants from taking unwanted breaks during the switch period, we constrained the switching period to 2 seconds. Trials where participants did not switch tasks within two seconds were not included in the analyses.

The task for the participants was to interpret instructions, and to apply these instructions to pairs of noun word stimuli. The pairs of noun word stimuli were pseudo-randomly picked with the constraint that two consecutive pairs could not use the same noun word(s). The instructions were novel in each trial but retained the same structure and logic. Each instruction asked participants to compare a pair of noun words on the basis of their attributes following a logic rule, and respond according to a response rule (see Table 1). For example, if the instruction were "SAME (for the logic rule), SWEET (for the attribute rule), LEFT-INDEX (for the response rule)", participants had to interpret this instruction as: "if the answer to 'is it sweet?' is the same for both noun words, I respond with my left-index finger. Alternatively, I respond with my left-middle finger." If the noun word stimuli were 'apple' and 'pear', the correct answer would be that both are sweet and therefore a correct response would be to press with the left-index finger ('x' key). Alternatively, if the word stimuli were 'apple' and 'salt', correct reasoning would involve realizing that one of them is not sweet, and therefore a correct response would be to press with the left-middle finger ('z' key). Note that 'true' responses were always given with a key press corresponding to the finger in the response rule. 'False' responses were always given with a key press associated with the other finger on the same hand. Participants had no time limit to respond. Immediately after response, an inter-stimulus interval (ISI) was presented for 1500 ms, indicated by a "+" at the center of the window. At the same time, auditory feedback was provided similar to the digit parity task. The trial ended after three stimulus-ISI sequences (see Figure 5.2).

5.2.1.3 *Instructions and noun words stimulus set*

In this experiment we used an adapted and translated (to Dutch) version¹ of the RITL

¹ The original RITL task instruction set as used by Cole et al. (2017) was in English. To make sure that fluency in language did not influence our results, we translated the rules to Dutch. Also, we decided to translate the rule 'niet tweede' (not second) as 'ontkennen tweede' (negate second). Pilot testing showed that participants were often confused by the 'not second' rule. This rule asks participants to judge whether the second word stimulus is e.g., not sweet. However, many participants interpreted this rule as having to judge whether the first word stimulus was sweet or not. Hence, we decided to translate this rule as the less ambiguous 'negate-second'.

Attribute rules	
Sweet	
Soft	
Green	
Loud	

Response rules	Key
Left-index	x
Left-middle	z
Right-index	n
Right-middle	m

Logic rules	Rule (first word + second word)
Same	yes + yes = true
	no + no = true
	yes + no = false
	no + yes = false
Different	yes + yes = false
	no + no = false
	yes + no = true
	no + yes = true
Second	___ + yes = true
	___ + no = false
Negate-second	___ + yes = false
	___ + no = yes

Table 5.1. Overview of all the different rules used to create the instructions for the RITL task. (left) List of attribute rules; (middle) list of response rules with associated keys; (right) list of logic rules with corresponding answering logic. Each instruction was compiled of a logic rule, an attribute rule, and a response rule (e.g., SAME, LOUD, RIGHT-INDEX). This meant that we had a total of $4 \times 4 \times 4 = 64$ possible novel instructions. All the instructions in the experiment were in Dutch but translated to English in this table to enhance understanding.

task instruction set from Cole et al. (2017). Each instruction consisted of three rules: a logic rule, a semantic rule, and a response rule. Our task set included four logic rules (same, different, second, negate-second), four attribute rules (sweet, soft, loud, green), and four response rules (left-index, left-middle, right-index, right-middle), resulting in $4 \times 4 \times 4 = 64$ unique instructions (see Table 5.1).

We used a normalized stimulus set of 64 Dutch noun words for the RITL task, including

16 noun word stimuli per attribute category. Each noun word in this set was positively matched on one attribute category and negatively matched on another category. For example, the word ‘Alarm’ positively matched on loud, but negatively matched on sweet.

To check for ambiguity in the attribute categories, we conducted a survey on an independent sample prior to testing ($N = 16$, 8 female; $M_{\text{age}} = 32.25$, $SD_{\text{age}} = 16.47$). This survey asked participants to report how much they agreed with a word stimulus fitting in a corresponding attribute category. Responses were made on a seven-point Likert scale. We included a noun word in the stimulus set when the average score on the positively matched category was greater than or equal to five (out of seven), and the score on the negatively matched category lower than or equal to three (out of seven).

5.2.1.4 Apparatus and set-up

We tested the participants individually in a dimly lit, windowless room. Participants were seated in front of a desk on which a display computer, monitor, eye-tracker, and head-mount was located. Eye tracking recordings of the left eye were performed with an EyeLink Portable Duo eye tracker from SR Research. This eye tracker is capable of measuring eye movements at a spatial resolution of 0.01 degrees, with an average accuracy down to 0.15 degrees. We used a sample rate of 500 Hz. The experiment was interfaced on a Mac mini running Windows 7, and presented on a 20-inch LCD monitor (1600 x 1200 pixels). The experiment was build using the Open Sesame experiment builder software (version 3.2.5). The code for the experiment is available online at the Open Science Framework (OSF; link to project: <https://osf.io/jx9ap/>). Before testing, we performed a 9-point calibration and separate validation using the software of the eye tracker. Drift check was performed at the start of each block to check the accuracy of the current calibration. We re-calibrated the eye-tracker before the start of a block when the participant removed their head from the head-mount during the block breaks.

5.2.1.5 Procedure

When participants entered the lab, they first received written instructions of the tasks in the experiment. We did not specifically instruct the participants that they could or should prepare for the RITL task. Instead, we only mentioned that the instruction for the RITL task was always visible. After reading the instructions, the participants signed an informed consent form and were asked to sit down in front of the computer and eye-tracker. We first adjusted the head-rest to the participant and adjusted the angle of the eye-tracker when necessary. Thereafter, we followed a calibration and validation procedure of the eye-tracker.

The experiment started with a practice phase for each of the two tasks. In contrast to

the testing phase, participants only performed one task in each trial during the practice phase. We decided to follow this procedure because the RITL task requires some practice to get familiar with the task, while the digit parity task does not. The tasks were presented at the center of the screen. Therefore, there was no windowed layout of the screen during practice.

In the first part of the practice phase, participants performed one block of trials for each version of the digit parity task. Participants started with one block of three trials with no working memory load, followed by one block of three trials with low load, and one block of five trials with high load. A recap of the instructions for each specific version was provided at the start of a block.

The second practice part included one block of 24 trials of the RITL task. In order to maximize practice with different instructions, practice trials included five noun word pair stimuli instead of three. The instructions during the practice phase were randomly picked from a practice set. This practice set contained four instructions, one for each combination of the four logic, attribute, and response rules. This practice set was randomly generated for each individual participant. A short recap of the task instructions of the different rules was provided before the start of the block. After finishing the second part the practice phase was finished. There was no eye-tracking during the practice phase.

During the testing phase, the participants performed six blocks of eight trials (48 trials in total). Only one version of the digit parity task was performed in each block, resulting in two blocks for each condition (16 trials per condition in total). The order of blocks and conditions was randomized by participant. For each block of trials, we picked eight instructions without replacement from the instruction set with the constraint that each specific logic rule should occur equally often in the block. This ensured that there could not be any differences in difficulty between blocks due to differential difficulty of the logic rules. The order of presentation of the instructions was also pseudo-random (without replacement), with the constraint that the same logic rule could not be repeated within two trials. Participants were allowed to take a short break (~ 2 - 5 minutes) between the blocks.

After finishing the experiment, the participants were asked to fill out a short paper survey with questions regarding their age, educational level, handedness, bilingualism, and multitasking strategy (specifically, “did you prepare for the RITL task?”, and “was this different for the different difficulty levels of the digit parity task?”). After completion, written and verbal debriefing was provided to the participant. The total duration of the experiment was approximately 1.5 hours, divided between 15 minutes for instructions

and setting up of the eye tracker, 20 minutes for practice, 40 minutes for testing, and 10 minutes for the survey and debriefing.

5.2.1.6 *Data analysis*

5.2.1.6.1 *Preprocessing*

Before analyzing the data, we first performed several pre-processing steps. Blinks that were detected by the eye tracker software were removed from the data, including 100 milliseconds before and after the event. This resulted in the removal of 6.08 % of the eye tracking recordings, ranging from 0.3 to 15.3 % across participants. Thereafter, we checked for trials in which participants did not perform a task switch within 2 seconds after completion of the digit parity task. As a result, we removed 5.0 % of all the trials (range 0.0 – 70.8 % across participants). Lastly, we examined QQ-plots of the response times in the digit parity task and RITL task for outliers. We decided to exclude trials with response times below 250 milliseconds and above 12 seconds for the RITL task (i.e., > 2 SD), which corresponded to 1.5 % of all observations.

Following artefact and outlier correction, we examined the eye tracker data for indications of preparation for the upcoming RITL task. First, we extracted all the fixations from the data (i.e., removing data belonging to saccades). Thereafter, we determined whether fixations were located in the window of the digit parity task, the window of the RITL task, or the RITL instruction (200x200 pixel area around the instruction). We defined the start of a preparation attempt as a switch in gaze position from the window of the digit parity task to the instruction of the RITL. In other words, a preparation attempt was completed when a fixation in the digit parity task window was followed by a fixation on the instruction. We defined the end of a preparation attempt as the moment when the participants' gaze returned in the window of the digit parity task. For every participant, trial, and sub-trial (a sub-trial is defined as a stimulus + ISI sequence within a single trial of the digit parity task, each trial had 10 sub-trials), we determined whether a preparation attempt was committed and if it took place, what was its duration in milliseconds.

5.2.1.6.2 *Statistical analysis*

We used linear-mixed effects models to assess statistical significance of effects. Response times from both tasks were fitted assuming a gaussian distribution. Since response times were skewed to the right – and therefore did not strictly meet the assumption of gaussian data – we transformed the observations using the natural logarithm. Error rates in both tasks and preparation attempts were expressed as binomial variables. We coded incorrect responses as 1 and correct responses as 0. Preparation attempts were considered on a trial and sub-trial level. On a trial level, we coded preparation attempt as 1 when at least one preparation attempt was observed during the trial of the digit

parity task, otherwise it was coded as 0. For sub-trials we checked whether a preparation attempt was observed during the ten stimulus-ISI (i.e., a sub-trial) sequences of the digit parity task. Both error rates and preparation attempts were fitted with binomial linear-mixed effects models.

We used a series of log-likelihood-based step-wise backward model comparisons to find the models that best fitted our data, which proceeded as follows. For every model, we first determined the random effects structure. We started with a model including the dependent variable of interest, main effects for the predictors, and a maximized random effects structure (Barr et al., 2013). Leaving out one term at every step, we kept the simpler model that had the most parsimonious structure without sacrificing explained variance (Matuschek et al., 2017). Models that did not converge or provided singular fit were not considered. In addition, we checked for potential overfitting problems in the random effects model, using a principal component analysis of the random effect variance-covariance estimates (Bates et al., 2015). After determining the random effects structure, we fitted a full model including all possible two-way interactions and the random effects. We then proceeded to remove fixed effects until the best model was found using a similar backward model comparisons procedure.

The analysis was performed in R (version 3.5.0). Gaussian and binomial linear-mixed effects models were fitted using the ‘lmer’ and ‘glmer’ function, respectively, from the ‘lme4’ package (version 1.1-19; Bates, Maechler, Bolker, & Walker, 2018). Principal component analysis on the random effects structure was performed with the ‘rePCA’ function from the same package. Statistical significance was determined using a significance threshold of $\alpha < 0.05$. Unlike the binomial linear-mixed effects models, the gaussian linear-models did not provide p-values. For these models, $|t| > 2$ was used as significance threshold. We have made the raw data, preprocessing scripts, and a markdown script of the performed analyses available online (see OSF project: <https://osf.io/jx9ap/>).

5.2.2 Results

We report the results from the final models that were determined by a backwards model fitting procedure. In short, this procedure helps to determine the best-fitting and parsimonious model by reducing model complexity step by step until model fit is harmed (more detailed explained in *Methods: Statistical Analysis*). To improve the readability of the section, we decided not to mention the configurations and summary tables of the final models in-text but leave those in a markdown script of the performed analysis available in the OSF project online. We excluded one participant from all the performed analyses below. This participant did not switch between the two tasks within two seconds for the majority of the trials (71 %).

Figure 5.3. Overview of the performance measures (error rate and response times) for the digit parity task and RITL task. The mean error rate (proportion) was calculated by averaging the subject means of the binomial error rate for each working memory load condition in the digit parity task (A) and for each trial set in the RITL task (C). The mean response times in (B) and (D) reflect the mean of the subject means. Error bars represent one standard error of the subject mean.

5.2.2.1 Behavioral analysis

We will first examine how participants performed in the digit parity task and rapid instructed task learning (RITL) task. Figure 5.3 shows the error rates and response times in both tasks.

5.2.2.1.1 Digit parity task

As expected, we found that working memory load in the digit parity task had a significant influence on error rate ($\chi^2(2) = 104.69, p < .001$). Participants made more errors with

increasing working memory load (see Figure 5.3A). Interestingly, we did not see the same pattern for response times. As can be seen in Figure 5.3B, the response times appear to be very similar across conditions. Our analysis supported this observation. Participants were significantly quicker in the low-load ($\beta = -77.40$ ms, $t = -4.66$) and high-load condition ($\beta = -94.59$ ms, $t = -4.38$) compared to the no-load condition, but this difference was only small.

5.2.2.1.2 *Rapid instructed task learning (RITL) task*

Consistent with the idea that participants learn to apply the instruction to the noun words in the RITL task, we observed that the first set of noun word stimuli had the highest error rates and slowest response times (see Figure 5.3C&D). Error rates significantly decreased for the second ($\beta = -4.44$ %, $z = -5.16$, $p < .001$) and third set ($\beta = -3.06$ %, $z = -3.42$, $p < .001$) compared to the first. Similarly, response times were faster on the second ($\beta = -2898.31$ ms, $t = -67.86$) and third set ($\beta = -2998.02$ ms, $t = -71.98$). Since response times were only measured from the moment that participants had switched tasks, the decrease in response time from the first to the second set likely reflects the time needed to interpret the instruction (difference in observed means = -3129.36 ms). The difference found in this study agrees with existing research showing that preparation for the RITL task can take multiple seconds (3503 ms in Cole, Patrick, Meiran, & Braver, 2017).

5.2.2.2 *Eye tracking analysis*

We used eye tracking to determine how working memory load in the digit parity task influenced preparation for the RITL. In addition, we assessed how preparation affected performance in the digit parity task and the upcoming RITL task for which the preparation was done.

5.2.2.2.1 *Individual differences in preparation frequency*

Since we did not specifically instruct participants to prepare for the RITL task, we expected that there would be individual differences in how often participants prepared. A preparation attempt was defined as an eye movement from the window of the digit parity task to the RITL instruction and back again (see *Methods: Preprocessing*). The results confirmed this hypothesis (see Figure 5.4). Overall, participants performed at least one attempt on 28.2 % of the trials (SD = 26.1 %). This average varied across the participants, ranging from 0.0 % (i.e., no preparation observed during experiment for a participant) to 89.6 %.

5.2.2.2.2 *Characteristics of preparation under working memory load*

We investigated the influence of working memory load in the digit parity task on preparation for the RITL by looking at the probability and frequency of eye movement

Figure 5.4. The proportion of trials in which at least one preparation attempt (fixation on the window of the RITL task) was counted for each participant.

consistent with preparation. The results are plotted in Figure 5.5.

We expected that participants would be less likely to engage in preparation when the digit parity task placed (more) demands on working memory, due to interference with the planning process. In line with this hypothesis, the probability of preparing for the RITL task decreased when the digit parity task required working memory ($\chi^2(2) = 164.17$, $p < .001$) compared to the no working memory condition. In addition to affecting the likelihood of preparation, we also found that working memory load affected the number of preparation attempts within a trial ($\chi^2(2) = 28.43$, $p < .001$). In trials where we observed at least one preparation attempt, we found a lower number of attempts for low-load trials (estimated adjustment in number $\beta = -0.54$, $z = -4.03$, $p < .001$) and high-load trials ($\beta = -0.66$, $z = -4.40$, $p < .001$) compared to trials without working memory load (intercept $\beta = 1.95$, $z = 8.49$, $p < .001$). The difference between low- and high-load trials was not significant ($\beta = -0.13$, $z = -0.88$, $p = .38$).

5.2.2.2.2 The moment of preparation

The results showed that current working memory load affected preparation in a trial. Here, we examined the moment of preparation. Figure 5.6 shows the proportion of preparation attempts across the sub-trials of the digit parity task for each condition. Recall that one sub-trial includes the period of one digit stimulus and the following inter-stimulus interval (~3 seconds in total). As can be seen from Figure 5.6, participants

Figure 5.5. Influence of working memory load in the digit parity task (i.e., no, low, or high) on several characteristics of preparation behavior: (A) the likelihood to prepare in a trial and (B) the frequency of preparation attempts in a trial. Measures on the y-axes for all figures reflect the average of subject means. Error bars represent one standard error of the subject mean.

prepared most at the start of no-load trials. This pattern is much less apparent for low- and high-load trials. Our results confirmed these observations. The fitted model predicted a smaller effect of sub-trial (i.e., time on trial) on the incidence of preparation (1 = did prepare in sub-trial; 0 = did not prepare in sub-trial) for low- (adjustment $\beta = +0.15$ (logit scale), $z = 4.96$, $p < .001$) and high-load trials (adjustment $\beta = +0.15$ (logit scale), $z = 4.30$, $p < .001$) compared to no-load ($\beta = -0.20$ (logit scale), $z = -11.12$, $p < .001$). In effect, the predicted slopes for low- and high-load trials ($\beta = -0.20 + 0.15 = -0.05$) were close to zero. Therefore, while participants prepared most often early in the trial, the likelihood of preparation was relatively constant over the course of the trial when the main task required working memory.

5.2.2.2.3 Benefit of preparation

Now we examine how demand on working memory in the digit parity task influenced the usefulness of preparation for the upcoming RITL task. In Figure 5.7, we plotted the distribution of the average error rate and response time in the RITL task across subjects, for trials with and without preparation, in each load condition.

We only analyzed performance (i.e., error rate and response time) on the first set of noun words. That is because you only need time to prepare for the RITL (i.e., interpret

Figure 5.6. The proportion of preparation attempts across sub trials of the digit parity task for each working memory load condition (i.e., no, low, and high). One sub trial reflects the sequence of a digit stimulus and inter-stimulus interval in the digit parity task (~3 seconds). The shaded area around each line represents one standard error of the subject mean.

the instruction) on the first set of noun word stimuli; after that you can re-use the built-up task representation for the remaining sets of stimuli. We also accounted for any influence of task switching time from the digit parity task to the RITL task. Exploratory analyses revealed that longer switching times were predictive of faster response times on the RITL task ($\beta = -0.11$ (log scale), $t = -4.06$), and that preparation during the digit parity task reduced switching time ($\beta = -0.19$ (log scale), $t = -7.09$). Quicker switching of tasks following preparation may indicate that participants were more ready for the RITL task, or at least, that they already had the second task in mind. On trials without preparation, participants apparently needed more time to ready themselves for the RITL task. Since we were here mainly interested in what participants did during the digit parity task, we accounted for the influence of switching time by including it as a predictor in our models.

We first looked at how preparation during the digit parity task influenced error rates in the RITL task. Although the RITL task was self-paced, and therefore potentially not very sensitive to improvements in error rate, we expected that preparation would decrease error rate in the RITL task. Nonetheless, our results showed no reliable main effect of preparation on the error rates in the RITL task ($\chi^2(1) = 3.16$, $p = .08$), nor did we find support for an interaction effect with the load conditions ($\chi^2(2) = 0.16$, $p = .92$). As can be seen in Figure 5.7A, most subject averages were concentrated around the same error

Figure 5.7. A violin plot showing the distribution of (subject) mean error rates (A) and response times (B) on the first trial set in the RITL task, separated by working memory load condition of the digit parity task (i.e., no, low, high) and the occurrence of a preparation attempt in a trial (yes = blue; no = white). The grand subject means are indicated with red cross bars. The black dots in the figure indicate the individual subject means. Error bars represent one standard error of the subject mean.

rate values for both preparation and non-preparation trials. Some subjects had an error rate of 1 during preparation trials, indicating that they always responded incorrectly when they prepared in advance. Post-hoc checking revealed that these participants only prepared on a few trials, hence, the observed error rate in these participants may be coincidental.

In contrast to error rates, we did find support for an interaction effect between preparation and the working memory load condition of the digit parity task on response times of the RITL task ($\chi^2(2) = 40.71, p < .001$). In line with our hypothesis that concurrent working memory load interferes with planning, the model predicted no significant influence of preparation on response times in low-load ($\beta = -136.92 \text{ ms } t = -0.49$) and high-load trials ($\beta = +406.72 \text{ ms } t = 1.29$). However, preparation did significantly predict faster response times for no-load trials ($\beta = -1014.64 \text{ ms}, t = -4.55$). It should be noted, however, that the distribution of response times extends to lower response time values in low-load trials with preparation, in comparison to trials without preparation. This indicates that, despite there not being an overall effect of preparation, some participants did benefit from preparation during low-load trials.

We expected that any reduction in response time in the RITL task would be the result of

Figure 5.8. Violin plots of the instruction gaze time (A) and stimulus gaze time (B) before response on the first trial set in the RITL task, separated by working memory load condition of the digit parity task (i.e., no, low, or high) and the occurrence of preparation in a trial (yes = blue; no = white). The instruction- and stimulus gaze durations were determined by calculating the total duration of eye fixations within a 200x200 pixel window around the noun word stimuli. The black dots indicate individual correct RITL trials of the first trial set. Grand subject means are indicated with the red cross bars.

needing less time to interpret the instruction. Therefore, we determined for each correct trial how long participants looked at the instruction and stimulus (i.e., the noun words) separately before they made their response. Figure 5.8 shows the distribution of the instruction gaze time (Figure 5.8A) and stimulus gaze time (Figure 5.8B) of these trials separated by the working memory load condition and preparation (yes/no).

5.2.2.2.4 Cost of preparation

Finally, we examined how preparation influenced performance on the digit parity task. We expected that preparation would increase error rates and response times on probes that were close to the onset of preparation attempts. Furthermore, we predicted that this increase would be larger in conditions with higher working memory load. To investigate this, we determined the performance on probes for which a preparation attempt was observed in the preceding sub-trial.

The results showed that preparation prior to a probe had no significant effect on error rate aside from the influence of working memory load (see Figure 5.9A; $\chi^2(2) = 0.37$, $p = .54$). Furthermore, we found no significant main effect of preparation on response times (see Figure 5.9B; $\chi^2(2) = 0.02$, $p = .88$). Contrary to our prediction, we also failed to find

Figure 5.9. Violin plots of the average error rate (A) and response time (B) on probe digits preceded by a preparation attempt (blue) or not (white). Individual black dots indicate subject averages. The red cross bars indicate the grand subject means. Error bars represent one standard error of the subject mean.

support for an interaction effect between whether there was a preparation attempt and working memory load on error rates ($\chi^2(2) = 2.41, p = 0.30$) and response times ($\chi^2(2) = 4.93, p = .09$).

5.2.3 Discussion

In Experiment 1, we examined how working memory load in an ongoing task affected how much and when participants engaged in preparation for an upcoming task. Furthermore, we were interested in how useful (for the upcoming task) and costly (for the current task) this preparation was. Based on the threaded cognition theory (Salvucci & Taatgen, 2008), we predicted that continuously maintaining digit items in working memory would be a bottleneck for preparation processing, thereby limiting the opportunities for preparation. Our results were in line with this prediction. We found that demand on working memory in the ongoing digit parity task decreased the likelihood of task preparation and lowered the number of preparation attempts.

Along with limiting opportunities to engage in preparation, threaded cognition theory also predicted that the benefit of preparation to the RITL task would diminish when the ongoing task placed demands on working memory. That is because working memory cannot be shared between the ongoing task and preparation processes when those occur at the same time. Our results were in line with this prediction. First of all, we

found that preparation improved response times (but not error rate) for the upcoming RITL task when the ongoing digit parity task required no working memory. However, no support for such an improvement was found in low- and high-load trials. Moreover, we found that preparation sometimes resulted in not needing the instruction anymore during the RITL task. The proportion of trials for which this was the case decreased with increasing working memory load. Therefore, it seems likely that we can effectively build a task representation while performing an ongoing activity, assuming that the ongoing activity does not need working memory. All in all, we conclude that our results show that concurrently maintaining items in working memory interferes with preparing for a second task.

Interestingly, we found no support for a cost of preparation to the digit parity. This does not necessarily mean that preparation caused little interference in the digit parity task. After all, the previous results indicated that working memory load in the digit parity task did interfere with preparing for the RITL. Therefore, the lack of a clear cost of preparation may indicate that participants were not willing to sacrifice performance on the digit parity task in favor of preparing for the RITL.

To conclude, our first experiment showed that it is possible to prepare for an upcoming task during an ongoing task with little to no cost. In addition, current working memory demand influences our willingness to prepare, and working memory load diminishes the usefulness of preparation.

5.2.3.1 *Limitations, alternative strategies, and alternative explanations*

A notable outcome and limitation of this experiment was the large variation in preparation across the participants. In fact, almost half of the participants (17 out of 36) did not prepare at all or only very little (on less than 20 % of the trials). Although there are several possible reasons for not preparing, the bottom line is that participants have a choice: the experiment does not require them to prepare, so some elect not to. We therefore cannot assume that they are either more or less focused on the task. These participants may have used opportunities for preparation to engage in other thoughts, such as day dreaming or planning for other tasks (see Klinger & Miles Cox, 2011; Smallwood et al., 2011). This is supported by a post-hoc analysis that showed that the group of participants who spent little time preparing for the RITL (i.e., on less than 20 % of the trials) did not perform significantly different on the digit parity task compared to participants who did prepare for the RITL regularly (i.e., on more than 20 % of the trials; error rate: $\chi^2(1) = 2.86$, $p = .09$; response time: $\chi^2(1) = 0.14$, $p = .71$). Irrespective of the reason, the lower number of observations for preparation in comparison to non-preparation cases did limit our statistical power to measure how working memory load influenced preparation and its

usefulness. The mixed effects regression analysis that we performed was robust against such imbalances, because the effects of individual participants were nested in random effects (see e.g., Martin, Nussey, Wilson, & Réale, 2011). Nonetheless, replication of our results is warranted.

Another limitation in this experiment was that participants could use an alternative strategy in the low-load version of the digit parity task which enabled them to bypass the need for working memory. Instead of maintaining the digit in working memory, participants could have also prepared the response in advance by placing their finger on the key corresponding to the parity of the digit. This might explain why we see in Figure 5.7B that some participants did show improvements in response times following preparation in low-load trials.

An alternative explanation for the results in this experiment is that the task versions with working memory load were simply more difficult. After all, both task versions required more mental operations/steps for successful performance than the no-load task. An interesting question is then what would happen if the task was made to be easier than the current no-load condition. For example, we could remove the parity decision and therefore only require participants to indicate whether they see a probe (i.e., traditional o-back task). If the reduction in preparation were the result of (increasing) task difficulty, one would predict that the easier condition would increase the likelihood of preparation as well as enhance the usefulness of preparation beyond the current no-load condition. However, threaded cognition theory would not predict this. Although the likelihood of preparation may increase a little bit as a result of reduced time on task, the usefulness of preparation should remain the same. That is because only demand on working memory is predicted to directly interfere with building the task representation. We decided to conduct a second experiment to test these predictions and to address the aforementioned limitations.

5.3 EXPERIMENT 2

There were several changes in Experiment 2 compared to Experiment 1. Because we wanted to disentangle task difficulty from working memory load, we added a new ‘control’ version to the digit parity task. This new version was designed to be easier than the no-load version of the previous experiment. In this new version, participants simply responded with a key press whenever a probe digit was presented on the screen. Therefore, the only difference with the no-load version was that participants did not need to judge the parity of the digit. To investigate the influence of task difficulty and demand on working memory, we contrasted this control version with the no-load version and low-load version. If task difficulty was the main reason why preparation became less useful

in the first experiment, we should see a step-wise decrease in usefulness of preparation from the control to the low-load version. However, if a bottleneck in working memory was the primary reason for the effect, we should only see a decrease in benefit from preparation for the low-load version, but not between the control and no-load version.

We decided to adapt the low-load version in this experiment to account for a potential confound in the previous experiment. Namely, participants could potentially bypass the need of working memory by preparing the response in advance. Therefore, we replaced the low load version with a 1-back task in Experiment 2, in which participants needed to report whether the probe digit is the same or different from the previous digit. We decided not to keep the high-load version in the second experiment to prevent the experiment from becoming too long.

Apart from changes to the experimental paradigm, we also adapted the instruction about the possibility of preparation from the first experiment. The instruction for Experiment 1 only mentioned that the RITL instruction task was always visible in the right upper corner. We adapted this instruction to include that this also creates the opportunity for preparation, and that preparation may be beneficial to performance. With this adapted instruction, we hoped to motivate more participants to engage in preparation attempts for the RITL task.

5.3.1 Methods

5.3.1.1 Participants

We recruited in total 37 participants (17 female; $M_{\text{age}} = 21.19$; $\text{range}_{\text{age}} = 18 - 26$) from the University of Groningen and the Hanze University of Applied Sciences. We adopted the same entry requirements and monetary compensation for the participants as Experiment 1.

5.3.1.2 Experimental paradigm

The experimental paradigm in Experiment 2 was only different from Experiment 1 in the digit parity task. The instruction and noun word set were the same as in Experiment 1. Figure 5.10 gives an overview of the full experimental paradigm.

Similar to the first experiment, we included three versions of the digit parity task. The control version asked participants to press the ‘m’ key on the keyboard as soon as a probe digit appeared on the screen. Hence, the control version only required visual attention to the stimuli. The second version was equivalent to the no-load version of the previous experiment. Therefore, in addition to recognizing the probe digits the participants responded corresponding to the number parity of the digit. They pressed ‘m’ for even and ‘n’ for odd. We refer to this version here as ‘CRT’ (i.e., choice-reaction

Figure 5.10. Overview of a single trial in Experiment 2. The paradigm was only different from Experiment 1 in the digit parity task (upper part in both figures). The digit parity task had three versions in Experiment 2: a control version that only required recognizing the probe digit (control), one version that added the need to retrieve number parity from memory (CRT), and one version placing demand on working memory (1-back). Only one version was performed in each block of trials. Text in the experiment was in Dutch, but translated in this figure to English to facilitate understanding.

time) to emphasize its difference from the control version. The third and last version placed demand on working memory. Similar to the low-load version in the first experiment, participants continuously remembered one digit, and compared this digit with the probe. The participants were asked to press the ‘m’ key when the probe digit was the same number as the previous digit. Participants pressed the ‘n’ key when the probe digit was dissimilar from the previous digit. The number of matching and non-matching probe digits was random on each trial. We refer to this third version as 1-back since it is equivalent to a traditional digit 1-back task.

5.3.1.3 Apparatus and set-up

The eye tracker set-up was the same for Experiment 2. The code for the experiment can be found in the OSF project online.

5.3.1.4 Procedure

The procedure slightly differed from the first experiment in the provided instructions, the first part of practice phase, and the block order in the test phase.

In the hope of motivating more participants to engage in preparation we adapted the instruction about the possibility of planning. The new instruction read (italics were new for Experiment 2): “The instructions are visible in the upper right corner for Task B [RITL task] during the entire trial. *This provides the opportunity to prepare for Task B. This may improve your performance* (translated from Dutch).” This instruction was included in the written instructions before the start of the experiment, and on the screen before the start of the testing phase.

The first part of the practice phase now consisted of the control condition, followed by the CRT and 1-back condition. The second part of the practice phase was similar to the previous experiment.

For the testing phase we used a pseudo-random order for the blocks and conditions instead of the full random order in the first experiment. Participants performed each version of the digit parity task in one the first three blocks and in one the last three blocks. The order of conditions in the first and second half was random.

5.3.1.5 Data analysis

5.3.1.5.1 Preprocessing

Blink and artefact rejection resulted in the removal of 5.3 % of eye-tracking recordings, ranging from 1.1 to 23.4 % across participants. In addition, we checked for trials in which participants did not perform a task switch within 2 seconds after completion of the

digit parity task. We removed the observations of 4.5 % of the trials, (range = 0.0 - 54.2 % across participants). Lastly, we removed response time observations for the RITL task that were below 250 milliseconds or above 12 seconds. This corresponded to 0.8 % of all the observations.

5.3.1.5.2 Statistical analysis

We performed the same statistical analysis for Experiment 2 as for Experiment 1.

5.3.2 Results

Similar to Experiment 1, we report the results of the final models that were determined by backwards model fitting. Full details on the configuration of the models can be found in the markdown script of the performed analysis in the OSF project online. We excluded two participants from all the performed analyses below. One participant did not switch between the two tasks in time for the majority of the trials (54%). The other participant aborted the experiment during the practice phase, hence, we had no usable data for this participant.

5.3.2.1 Behavioral analysis

Figure 5.11 shows the observed average error rates and response times in both tasks. To facilitate comparison with Experiment 1, we marked the averages of the previous experiment in the graphs for the overlapping conditions.

5.3.2.1.1 Digit parity task

As expected, participants made very few errors when they only needed to respond to the presence of a probe digit (i.e., control version; see Figure 11A). Any errors that were made in this condition reflect missed responses or accidental incorrect button presses (e.g., 'n' instead of the correct 'm' key). The number of errors was higher when participants also needed to judge the parity of the probe digits (i.e., CRT version; estimated adjustment $\beta = +1.92\%$, $z = -5.90$, $p < .001$) and when they had to judge the similarity with a prior digit (i.e., 1-back version; estimated adjustment $\beta = +9.34\%$, $z = -17.07$, $p < .001$). The difference in error rate between the 1-back and CRT version was significant (estimated adjustment $\beta = +7.42\%$, $z = 11.47$, $p < .001$), indicating that more errors were made in the 1-back version compared to the CRT version. Similar patterns were found for response times. Participants were quickest in the control version (predicted $M = 500.90$ ms; reflected in intercept $\beta = 500.90$ ms, $t = 251.69$) and slower in the CRT (estimated adjustment $\beta = +249.21$ ms, $t = 24.83$) and 1-back version (estimated adjustment $\beta = +262.62$ ms, $t = 22.82$). The difference in response time between the CRT and 1-back version, however, was not significant (estimated difference $\beta = 13.42$ ms, $t = 1.11$).

Figure 5.11. Task performance for the digit parity task and RITL task. The cross marks in the graph reflect the corresponding averages from Experiment 1. Error bars represent one standard error of the subject mean.

A few things stand out compared to Experiment 1. First of all, participants made more errors in the 1-back version compared to the corresponding low-load version of the first experiment. We also do not see the drop in response time anymore from the CRT to 1-back task. Together, this suggests that we were successful in improving the manipulation of working memory load in this version. We also found that the total duration of preparation was lower in CRT compared to control trials ($\beta = -1167.13$ ms, $t = -6.10$). The difference in total duration between 1-back and CRT trials was significant ($\beta = -308.79$ ms, $t = -2.74$).

5.3.2.1.2 Rapid instructed task learning (RITL) task

The general pattern in error rate and response time in this experiment was very similar

Figure 5.12. Proportion of trials in which we observed at least one preparation attempt for each participant in the second experiment. Preparation attempts were determined using the same method as in the first experiment.

to Experiment 1 (see Figure 5.11B & C). Compared to the first trial set, participants made fewer errors in the second ($\beta = -2.55\%$, $z = -2.55$, $p = .011$) and third set ($\beta = -3.61\%$, $z = -3.72$, $p < .001$). Similarly, participants were quicker in the second ($\beta = -1966.85$ ms, $t = -47.69$) and third set ($\beta = -2055.17$ ms, $t = -51.52$). Noticeable is that participants overall made more errors in this experiment compared to the first. This increase in errors is unlikely to be caused by a shift in speed/accuracy trade-off because the response times for the second and third set were very similar to the first experiment (second = +81.24 ms; third = +15.39 ms). Interesting is that participants in the current experiment were substantially faster in the first set compared to the second (difference = -883.82 ms). These faster response times may reflect the effect of improved preparation in the current experiment since preparation attempts were more common (see below).

5.3.2.2 Eye tracking analysis

5.3.2.2.1 Individual differences in preparation frequency

Figure 5.12 shows the proportion of trials for which we observed at least one preparation attempt for each participant. Preparation attempts were determined in the same way as for Experiment 1. In Experiment 2, preparation was much more common (66.2 % compared to 28.2 % in Experiment 1). This suggests our adapted instruction worked.

Figure 5.13. Overview of the influence of task version in the digit parity task (i.e., control, CRT, or 1-back) on several characteristics of preparation behavior: (A) the likelihood of preparation in a trial and (B) the number of preparation attempts in a trial. Measures on the y-axes for all figures reflect the average of subject means. The cross marks in the graph reflect the corresponding averages from the first experiment. Error bars represent one standard error of the subject mean.

5.3.2.2.1 Characteristics of preparation

Figure 5.13 shows how the probability (Figure 5.13A) and frequency (Figure 5.13B) of preparation differed between the versions of the digit task. As we expected, the likelihood of preparation for the RITL task was different depending on what version of the digit parity task was performed ($\chi^2(2) = 135.76$, $p < .001$). The highest probability of preparation was found for the control version (intercept $\beta = 93.01\%$, $z = 4.861$, $p < .001$). Contrary to our expectation, the probability of preparing was lower on CRT trials compared to control ($\beta = -5.79\%$, $z = -3.57$, $p < .001$). However, with a predicted decrease of 5.79%, this difference in probability was only very small. In contrast, our results showed substantial decrease for 1-back trials compared to control trials ($\beta = -31.77\%$, $z = -10.64$, $p < .001$) as well as CRT trials ($\beta = -25.98\%$, $z = -7.69$, $p < .001$). This replicates Experiment 1 in showing that less working memory load decreases the probability of preparing. We also noticed that preparation in the 1-back task was much more common in this experiment than in Experiment 1 (observed difference = 28.26%). This may indicate that participants were more motivated to prepare in the low working memory load condition in this experiment compared to Experiment 1.

Apart from the probability of preparation, we also found that the number of preparation attempts within a single trial differed between task versions ($\chi^2(2) = 55.28$, $p < .001$).

Figure 5.14. The proportion of preparation attempts across sub trials of the digit parity task for each version of the digit parity task (i.e., control, CRT, and 1-back). One sub-trial reflects the sequence of a digit stimulus and inter-stimulus interval in the digit parity task (~3 seconds). The shaded area around each line represents one standard error of the subject mean.

Contrary to our expectation, our results showed a decrease in the number of preparation attempts for CRT trials compared to control (estimated decrease in number $\beta = -0.55$, $z = -5.28$, $p < .001$). We also found a significantly lower number of preparation attempts in 1-back compared to CRT trials, replicating Experiment 1 ($\beta = -0.21$, $z = -2.06$, $p = .04$). However, this predicted difference was larger in the previous experiment (Experiment 1: $\beta = -0.54$).

5.3.2.2.2 Moment of preparation

Figure 5.14 shows the average proportion of preparation attempts across sub-trials for each version of the digit parity task. Similar to Experiment 1, we found that the proportion of attempts was highest early in the trial when no working memory was required by the task. From the first trial onwards, the likelihood of preparation attempts decreased. The results showed that the decrease preparation attempts over time was a bit smaller in CRT trials compared to control trials (adjustment in $\beta = +0.04$ (logit scale), $z = -2.15$, $p = .03$). Similar to the Experiment 1, the effect of time in 1-back trials was small ($\beta = -0.10$ (logit scale), $z = -7.01$, $p < .001$).

5.3.2.2.3 Benefit of preparation

We then examined the influence of the condition in the digit parity task on the

Figure 5.15. A violin plot showing the distribution of the (subject-)average error rates (A) and response times (B) on the first trial set in the RITL task, separated by the version of the digit parity task (i.e., control, CRT, 1-back) and the occurrence of a preparation attempt in a trial (yes = blue; no = white). Individual dots in the figure reflect individual subject means. The red cross bars indicate the grand subject means. The black cross marks in the graph reflect the corresponding averages from Experiment 1. Error bars represent one standard error of the subject mean.

usefulness of preparation. We expected no difference in the usefulness of preparation between control and CRT trials. Even though the control trials were easier, the additional demand on memory in CRT trials should not have interfered with preparation, because preparation depends on working memory resources. Therefore, we expected reduced benefit from preparation in 1-back trials in which the 1-back task competes for working memory resources with the preparation process.

Similar to Experiment 1, we found that participants switched more quickly between the tasks when they had prepared for the RITL ($\beta = -0.35$ (log-scale), $t = -9.29$). However, we did not find that task-switching time was predictive of quicker response times in the RITL ($\beta = -0.05$ (log-scale), $t = -1.77$). Therefore, we decided not to include task-switching time in the models below.

Figure 5.15 shows the distribution of (subject-)average error rates (Figure 5.15A) and response times (Figure 5.15B) for trials with and without preparation in each task condition. Similar to Experiment 1, we found no main effect of preparation on error rates ($\chi^2(1) = 1.21$, $p = .27$), nor did we find support for an interaction effect between preparation and the working memory load condition of the digit parity task on error

Figure 5.16. Violin plots of the instruction gaze time (A) and stimulus gaze time (B) before response on the first trial set in the RITL task, separated by the task version of the digit parity task (i.e., control, CRT, or 1-back) and the occurrence of preparation in a trial (yes = blue; no = white). The instruction- and stimulus gaze durations were determined by calculating the total duration of eye fixations within a 200x200 pixel window around the noun word stimuli. The black dots indicate individual correct RITL trials of the first trial set. Grand subject means are indicated with the red cross bars. The cross marks in the graph reflect the corresponding medians from Experiment 1.

rates ($\chi^2(1) = 3.04$, $p = .22$).

We did find a significant interaction effect between whether a participant prepared and the version of the digit parity task on response time ($\chi^2(2) = 31.65$, $p < .001$). In control trials, participants were estimated to be almost two seconds quicker when they had prepared for the RITL task ($\beta = -1758.74$ ms, $t = -7.46$) compared to when they had not. This estimated benefit was a bit smaller in CRT trials (adjustment $\beta = +376.70$ ms, $t = 2.23$), and substantially smaller in 1-back trials (adjustment $\beta = +1166.29$ ms, $t = 5.532$). In contrast to Experiment 1, the resulting estimated benefit of preparation in 1-back trials was significant ($\beta = -592.41$ ms, $t = 2.31$).

We determined for each correct trial how long participants looked at the instruction and stimulus for each task version of the digit parity task (Figure 5.16). We observed again no

Figure 5.17. Violin plots showing the average error rate (A) and response time (B) on probe digits preceded by a preparation attempt (blue) or not (white). Red cross bars indicate the grand subject means of Experiment 1, whereas the black cross marks indicate the grand averages of Experiment 1. Individual black dots reflect individual trials. Error bars represent one standard error of the subject mean. The cross marks in the graph reflect the corresponding averages from the first experiment.

influence of preparation ($\chi^2(1) = 0.12$, $p = 0.73$) nor of task version ($\chi^2(2) = 0.05$, $p = .98$) on stimulus gaze time. Therefore, Experiment 2 replicated our finding from Experiment 1 that preparation did not influence how much time participants needed to create a task representation for applying the instruction to the stimuli.

Similar to the previous experiment, we noticed here that participants frequently did not need to look at the instruction anymore after they had engaged in one or more preparation attempts for the RITL during the preceding tasks. The proportion of trials for which this was the case was very similar between control (28.2 % of trials) and CRT trials (25.9 % of trials), but noticeably lower in 1-back trials (13.4 % of trials). Non-parametric chi-square testing confirmed that the difference in proportion between control and CRT trials was not significant ($\chi^2(1) = 0.33$, $p = .57$). Furthermore, the proportion of trials with preparation attempts in the 1-back condition was significantly lower compared to the control ($\chi^2(1) = 16.46$, $p < .001$) and CRT condition ($\chi^2(1) = 11.90$, $p = .001$).

5.3.2.2.4 Cost of preparation

Lastly, we examined how preparation influenced performance on the digit parity task (Figure 5.17). Similar to Experiment 1, the results showed no support for an increase in error rate due to preparation ($\chi^2(2) = 0.05$, $p = .82$). We also found no support for an

interaction effect between preparation and the working memory load condition of the task on error rates ($\chi^2(2) = 3.21, p = .21$).

With respect to response times, the results showed no significant interaction effect between preparation and task version ($\chi^2(2) = 3.52, p = .17$), similar to Experiment 1. However, we did find support for a main effect of preparation on response times ($\chi^2(2) = 11.33, p < .001$). Participants were estimated to be 29 ms slower on the digit parity task immediately following a preparation attempt ($\beta = 28.98 \text{ ms}, t = 3.84$) compared to when they did not prepare.

5.3.3 Discussion

We conducted Experiment 2 to address two limitations of Experiment 1 and to replicate its core findings. One of the limitations of Experiment 1 was that a substantial group of participants (17 out of 36) prepared very little (i.e., < 20 % of all trials), leaving limited observations of preparation for our analyses. Here, we hoped to motivate more participants to prepare by adapting the experiment instructions. The results showed that we were successful in motivating more participants to prepare. Overall, participants prepared on the majority of all trials (66 %; compared to 28 % in Experiment 1). Only four out of 34 participants (12 %) prepared very little or not at all.

The other limitation we addressed in Experiment 2 was that participants could have bypassed the need for working memory by adopting a response preparation strategy in which they already placed the relevant finger on the button before the presentation of the next stimulus. To prevent this strategy, we replaced the 1-back version of the CRT in Experiment 1 (i.e., low-load condition) with a traditional 1-back task in Experiment 2. Although we do not have hard evidence that the 1-back task in Experiment 2 was a more reliable manipulation of working memory load, the observed increase in error rate and response time in comparison with Experiment 1 does indicate that participants experienced more difficulty with this task.

With Experiment 2 we were able to replicate most findings of Experiment 1. Similar to Experiment 1, we found that the extra demand on working memory in the 1-back task in Experiment 2 reduced preparation for the second task. Furthermore, we replicated the interference of working memory load in the first task on preparation for the second task. Preparation during 1-back trials resulted in a smaller improvement from preparation to response times in the RITL task compared to the control and CRT task that did not place demands on working memory.

The reduction in preparation and its effectiveness in the 1-back task brings us to our

second research question in Experiment 2: are the effects of the 1-back task on the extent and the usefulness of preparation the result of a bottleneck in working memory (as predicted by threaded cognition theory), or do they result from a general effect of task difficulty? Our results demonstrated a general effect of task difficulty, but also provided indications for a disproportionally larger effect of working memory load.

In line with threaded cognition theory, we observed that the probability of engaging in preparation attempts was very similar between the control and CRT task version. However, we also found support for a decrease in the number of preparation attempts in the CRT compared to control. This indicates that a seemingly simple additional process to identify number parity can already result in making participants less willing to disengage from their task.

The decrease in preparation translated in a decreased benefit for the task that was being prepared for. We observed that the benefit of preparation to response times in the RITL showed a step-wise decrease from control to 1-back. However, in contrast to what we should expect based on a general effect of task difficulty, the steps were not equally sized. The reduction in benefit to response times was much larger between the 1-back and control task (estimated adjustment of benefit: +1167 ms) compared with the CRT and the control task (estimated adjustment of benefit: + 377 ms). This suggests that preparation was disproportionally more difficult alongside a main task involving working memory, in line with the idea that there is interference between working memory resources needed for both tasks.

5.4 GENERAL DISCUSSION

This paper presented a paradigm to investigate how present task demands influence preparation efforts for an upcoming novel task. While busily engaged in a digit parity task, participants could make use of the opportunity to prepare for a rapid-instructed task learning (RITL) task that they knew was following. Specifically, they could inspect and interpret the task instructions in advance. In two experiments, we observed that many participants did indeed use this opportunity to prepare. While the digit parity task was still ongoing, participants glanced at the instruction of the upcoming RITL task. Such glances were not merely distractions, but likely reflected preparation attempts, since glances at the instruction were predictive of faster response times in the first trial of the RITL. Therefore, our research provided empirical evidence for the idea that people disengage from present activity in favor of a future task. In addition, our results indicate that such disengagements can be productive.

While some participants engaged in such productive distractions, a substantial group

of participants did not. In particular for Experiment 1, a large group of participants did not prepare – or only very little – across all conditions (i.e., $n = 19$ out of 36 prepared on less than 20 % of the trials). These individual differences reflect an essential property of the study that participants did not feel forced to prepare. Why some participants did prepare, while others only little, can have multiple reasons. For example, participants may not have (fully) registered, or misinterpreted, the instruction about the possibility of preparation. Also, some participants may have simply preferred not to multitask. What is clear is that the existence of an upcoming task was not enough to motivate all of them to prepare. When the task instruction emphasized the potential usefulness (i.e., incentive) of planning ahead in Experiment 2, more participants engaged in preparation.

5.4.1 When we prepare, and when it is useful

The research aim of this paper was to gain more understanding in how differential demands from a present task can influence whether and when we prepare, and to what extent preparation is useful.

5.4.1.1 Role of working memory load

We focused in particular on the potential influence of working memory load from the present task on task preparation. Assuming that working memory resources cannot be shared between the present task and preparation processes (see Borst, Taatgen, & van Rijn, 2010; Salvucci & Bogunovich, 2010), we predicted that participants would refrain from preparation when the present task required working memory. Furthermore, any preparation that would be committed under concurrent working memory load should be less effective, since the construction of the task representation in working memory cannot start or persist. Our results supported these predictions. In both experiments, we found that even maintaining a single item in working memory could already reduce the extent to which participants prepared and how beneficial it was to performance on the upcoming RITL task. Interestingly, we found no clear performance cost of preparation to the main digit parity task. Costs as a result of preparation also did not scale with the complexity of the digit parity task. Therefore, we conclude that our results demonstrate that working memory demands from the present task interfered with preparation. The lack of a clear cost to the present task shows that participants prioritized the two tasks well.

5.4.1.2 Influence of task difficulty

Working memory load was, however, not the only factor that influenced whether and how productively participants prepared. Experiment 2 demonstrated that increases in task difficulty alone, irrespective of working memory load, already reduced to what extent participants prepared and how useful their efforts were. Our experiments suggest

that there are at least two reasons why task difficulty influences whether or when we prepare. First increases in task difficulty may limit opportunities for preparation to occur. When the difficulty of the present digit parity task increased, also the number of mental operations needed to perform the task increased, leaving less time for task preparation. A second reason could be that participants may have already decided before the start of the trial how much to shift their focus to the future task. Such strategic shifts in multitasking have been found in other studies. For example, Janssen and Brumby (2015) examined whether people adopt an optimal strategy for interleaving two concurrent tasks when they attempt to maximize performance (see also Nijboer, Taatgen, Brands, Borst, & Van Rijn, 2013). Similar to our experiments, the researchers found that people altered their multitasking strategy in the face of changes in task difficulty. Additionally, they also showed that changes in incentives (see also Farmer, Janssen, Nguyen, & Brumby, 2018; Janssen et al., 2019), and individual differences in skill with the task(s), influenced how people decided to interleave two tasks. We did not examine the latter two variables, but they may be interesting to consider in future research.

5.4.2 Theoretical and practical contributions to the mind wandering, prospective memory, and multitasking literature

The experiments reported here brought together insights from mind wandering, prospective memory, and multitasking literatures. Each of these literatures emphasize different aspects of how we prepare in the context of ongoing activities. Mind wandering research studies when attention shifts from the current task to internal (planning) processes and what the self-reported content of these internal processes is (Seli, Konishi, Risko, & Smilek, 2018; Seli et al., 2016; Smallwood et al., 2009). Prospective memory research examines the encoding of intentions of future tasks and to what extent we can remember them across task activity (Einstein & McDaniel, 1996; Kliegel, McDaniel, & Einstein, 2000; McDaniel & Scullin, 2010; Scullin et al., 2018). Lastly, multitasking research provides theoretical frameworks that predict how preparation and current task processes interfere (Borst et al., 2010; Pashler, 1984; Salvucci & Taatgen, 2010; Wickens, 2002). Our experiments therefore contribute differently to each of these literatures.

With respect to mind wandering, the preparation attempts in this study show most resemblance to the deliberate and intentional forms in that literature. Currently, only a few studies have investigated how ongoing task demand influences intentional mind wandering. Seli and colleagues (2016) showed that the predictability of targets in a sustained attention task was associated with higher incidences of intentional mind wandering compared to a task context with unpredictable targets. More recently, another study demonstrated that a working memory load manipulation decreased intentional mind wandering (Seli, Konishi, et al., 2018). The present study combined

task demand manipulations with and without working memory load in one paradigm. Confirming prior work, we found that general increases in task demand as well as working memory load decreased intentional disengagements from the ongoing task. In addition, we found that introducing working memory load resulted in a disproportionately smaller preparation benefit (i.e., comparing CRT and 1-back version) compared to when only task demand without working memory load was manipulated (i.e., comparing control with CRT version). This demonstrated that the influence of working memory load on this was greater compared to general task difficulty. Hence, this study showed that the availability of working memory resources may have a strong influence on the decision of whether we disengage from a task to think about the future. Our study also provides a potential reason for this: preparing for the future under concurrent working memory load tends to be less effective.

We contributed to the prospective memory literature by studying how intentions for complex tasks are formed, maintained, and executed in the context of an ongoing task. Existing prospective memory research has already shown that people can maintain and execute complex intentions across delay activities (e.g., Kliegel, Mackinlay, & Jäger, 2008; Kliegel et al., 2000; Kliegel, McDaniel, Einstein, & Moor, 2007). Furthermore, it has demonstrated that concurrent working memory load can impair the encoding of simple future intentions (Einstein, Smith, McDaniel, & Shaw, 1997). What the present study adds is showing that it is possible to encode complex intentions during a task, which is necessary for performing the RITL task successfully (Cole, Braver, & Meiran, 2017). With our experiments, we have demonstrated that people can effectively engage in such working memory integration processes while concurrently performing another task.

Lastly, our research contributes to the multitasking literature in demonstrating that the predictions of multiple-resource theories of multitasking also generalize to voluntary task preparation during ongoing tasks. Multiple-resource theories, including threaded cognition, predict that we engage in a task as soon as the required resources are available (see e.g., Salvucci & Taatgen, 2008; Wickens, 2002). Additionally, these theories predict that we shield current task processes from other tasks when similar resources are required, otherwise, interference is observed. Indeed, we found that when resources for preparation were not occupied by the digit parity task, preparation occurred as soon as possible, i.e., at the start of the trial. Conversely, when resources were occupied, then participants tended to avoid task preparation, i.e., they shielded their current task processes.

While threaded cognition and other multiple-resource theories of multitasking would predict the observed results, there are also other multitasking theories that would

not have predicted our results. The threaded cognition theory falls in the middle of a spectrum of theories in how many bottlenecks there are in multitasking. Pashler (1984, 1994) advocated a single bottleneck in multitasking, namely during the period of preparing a response. The present results cannot be easily explained by this *single* bottleneck model. Because the single bottleneck model only assumes a bottleneck in central cognition, the predictions of this model are less refined than multiple bottleneck models which decompose central cognition. In Experiment 2, we found that interference from declarative memory (CRT task) on planning ahead was smaller compared to interference from working memory (2-back task). This difference in magnitude of interference cannot be easily explained with a single bottleneck account. On the other side of the spectrum there is Meyer and Kieras (1997)' *EPIC* theory, which suggests multitasking involves no bottlenecks (see also Kieras, 2017a, 2017b). According to their theory, interference between tasks occurs from inefficient scheduling of the tasks. This theory would have predicted no difference due to the additional working memory load in this study. In short, our study on concurrent task preparation is consistent with threaded cognition and other multiple-resource theories. However, our results cannot easily be explained with the single bottleneck model of Pashler, and are inconsistent with scheduling theories of multitasking.

5.4.3 Limitations and future research opportunities

While we were able to detect preparation attempts indirectly and showed its effects on task performance, we have no direct subjective judgments to support this interpretation of our data. The mind wandering literature has a long tradition of studying (spontaneous) thought-processes by incorporating periodic self-report (called 'thought probes') in the task (see Franklin, Smallwood, & Schooler, 2011; Kane et al., 2017; Robison, Miller, & Unsworth, 2019; Weinstein, 2017). Although self-report has its own limitations, it is worthwhile examining whether overt switching behavior also corresponds to self-reported preparation efforts.

Another possible limitation of this study lies in how participants were instructed about the possibility of preparation for the RITL during the digit parity task. In this study, we demonstrated two approaches. Experiment 1 used a subtle instruction mentioning only the availability of the instruction. Experiment 2, on the other hand, encouraged preparation more by adding phrases about the possibility of preparation and its potential benefits to performance. The differences in outcomes between the experiment were large, with a strong increase in preparation frequency for Experiment 2 compared to 1, demonstrating the influence of instruction on preparation behavior. However, it should be noted that it may not be possible to be fully neutral in the instruction. Not instructing at all may leave participants confused about the presence of the RITL instructions during

the digit parity task. On the other hand, explicitly instructing the participants to prepare may lead to unnatural behavior. Future research should explore in more detail the effect of motivation on preparation behavior.

Another interesting opportunity for future research is to further explore how task familiarity influences the interference created by task preparation. The RITL literature has proposed that while novel task preparation requires working memory integration to construct a task representation, preparation for practiced tasks can be done by simple retrieving its existing task representation from long-term memory (see e.g., Braverman & Meiran, 2010; Cole et al., 2010; Cole, Ito, & Braver, 2016; Cole et al., 2018; Mayr & Kliegl, 2000). Therefore, preparation for practiced tasks may not be as sensitive to interference from concurrent working memory load as preparation for novel tasks.

Another opportunity for future research is to examine individual differences in the tendency to prepare during ongoing tasks. Our research seems to support work on multitasking strategies that shows that making participants see the incentive of preparation can motivate them to prepare (see Farmer, Janssen, Nguyen, & Brumby, 2018; Janssen & Brumby, 2015; Janssen et al., 2019). It would be interesting to further explore the influence of incentives on whether and when participants prepare. Similarly, investigating the influence of individual differences in skill on the decision to prepare (or not) could provide interesting new insights in how people approach preparing for the future during a present task.

5.4.4 Conclusion

To conclude, in two experiments we observed that participants engaged in preparation for an upcoming task during ongoing activities. Such preparation was productive and induced only small costs to the current ongoing task. Broadly speaking there were two main factors that influenced whether and when participants engaged in preparation. First of all, we observed that there were large individual differences. While many participants did prepare for the RITL task, a substantial group did not or very little. For the group of participants that did prepare, the demand of the ongoing task was another major influence on whether and when they prepared. Both experiments showed that working memory load in the main digit parity task reduced preparation and compromised its usefulness. This suggests that working memory may play an important role in future task preparation. In addition to influence from working memory load we also found that general task difficulty reduced preparation and its usefulness. Task difficulty may have influenced preparation by limiting opportunities for it, as well as changing participants' multitasking strategy. Together, our study demonstrated that people can effectively prepare for the future while being busily engaged in a present task, and that it is possible

to track this preparation process empirically.

Chapter 6

Summary and conclusions

In this chapter we summarize the results of this thesis and discuss its practical implications.

SUMMARY AND CONCLUSIONS

This thesis focused on the topic of *self-generated thought*. Self-generated thoughts are thoughts that come up while already being engaged in another activity (Smallwood & Schooler, 2015). For example, our minds might wander off to concerns about the next day while casually watching series in the evening. We may even engage in self-generated thought during more demanding activities, such as reading this chapter (e.g., Franklin et al., 2011; Huijser et al., 2018; Reichle et al., 2010). Studies have estimated that we spent up to fifty percent of our waking lives in self-generated thinking (Kane et al., 2017; Killingsworth & Gilbert, 2010). Investing so much of our time in thought has consequences. Engaging in self-generated thought tends to distract us from other activities resulting in poor performance on those activities. Still, self-generated thought may also have adaptive properties (Mooneyham & Schooler, 2013). Engaging in self-generated thought may provide an opportunity to make sense of our experiences (Christoff et al., 2011), to come up with creative ideas (Baird et al., 2012; Leszczynski et al., 2017; Tan et al., 2015), to remind ourselves of due intentions (Szarras & Niedźwieńska, 2011), to plan ahead for the future (Baird et al., 2011; Huijser et al., 2021), and possibly much more – self-generated thought clearly plays a key role in our daily functioning.

The aim of this thesis is to provide more insight in the mechanisms that drive self-generated thought, as well as the consequences of self-generated thought for performance on ongoing and future activities. This chapter summarizes the key findings from this thesis, and discusses the practical implications of these results. The chapter is subdivided in three sections. The first section discusses how competition for cognitive resources influences self-generated thought. The second section focuses on how self-generated thought affects task performance and can be detected in eye tracking measures. The third and last section discusses the proposition that self-generated thought is an adaptive process. The chapter concludes with some practical implications.

6.1 Competition for cognitive resources

Currently, there are several theories that try to explain how self-generated thought can arise and/or persist during an ongoing activity (see e.g., Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016; Klinger & Miles Cox, 2011; McVay & Kane, 2009; Smallwood, 2013; Smallwood & Schooler, 2006). While these theories are certainly helpful for understanding the thought process, this thesis utilized a novel theoretical framework grounded in multitasking (see Salvucci & Taatgen, 2010), cognitive modeling (Taatgen, 2013), and goal competition (Altmann & Trafton, 2002; Gerjets, Scheiter, & Schorr, 2003) to investigate self-generated thought. The differentiating factor of this theoretical framework, and main reason for using it in this thesis, is that it provides precise (mechanistic) predictions about how self-generated thought can arise and persist.

The theoretical framework makes two propositions about self-generated thought (Taatgen et al., 2021; van Vugt et al., 2015). First of all, self-generated thought is an ordinary thought process relying on the same (or at least very similar) computations as any other thought process. And secondly, self-generated thought can occur in parallel to other thought processes. The framework predicts that self-generated thought can arise and persist when required cognitive resources are available for use. This is consistent with resource theories of self-generated thought, which claim that engaging in self-generated thought requires cognitive resources (Levinson et al., 2012; Smallwood & Schooler, 2006). In this thesis, we assumed that self-generated thought requires declarative memory resources for the retrieval of memories and working memory resources for temporary storage of the retrieved memories (Faber & Mills, 2018; Smallwood & Schooler, 2006; Spreng, Stevens, Chamberlain, Gilmore, & Schacter, 2010; Teasdale, Proctor, Lloyd, & Baddeley, 1993). Whether self-generated thought actually arises when declarative memory and working memory are available is dependent on an additional competition process. This competition happens between self-generated thought and the task at hand that wants to use the same resources. What process wins the competition depends on the relative priority of these processes at a given time. On the surface, this idea is consistent with the control failure theory of mind wandering (see McVay & Kane, 2009), which explains that mind wandering occurs when we fail to maintain task focus. The theoretical framework here, however, does not require any control processes to explain this, as the competition process is fully automatic. When resources for self-generated thought are available, processes from different tasks and goals will compete for these. The relative priority of these processes will determine what will be thought about next.

The priority of mental processes is in part determined by internal and external context. Internal context includes our currently active goals, retrieved memories, and working memory content. External context includes what we perceive with our senses, notably visual and auditory information. The influence of context on the priority of mental processes implicates that self-generated thought can be manipulated by processing distracting information from our internal and external context. In Chapter 2 we tested this prediction by means of investigating whether processing of self-referential material (e.g., ‘Am I kind?’) can leverage the competition process in favor of self-generated thought in a complex working memory task. We hypothesized that self-generated thought would compete with rehearsal for the task. Hence, processing self-referential material would lead to higher incidences of self-generated thought and, thereby, less rehearsal and poorer memory performance. Processing self-referential material may instigate self-generated thought because self-referential stimuli may evoke mental elaboration about the material (Conway & Pleydell-Pearce, 2000; Conway et al., 2004; Wessel et al., 2014).

Moreover, processing self-referential material may trigger mind wandering through spontaneous associations with e.g., personal goals and aspirations (e.g., Barnhofer et al., 2007; D'Argembeau et al., 2007; Klinger & Cox, 2004). The results showed that self-referential processing indeed increased reports of mental elaboration about the self-referential material. Furthermore, we found a small increase in mind wandering following self-referential processing. The increase in self-generated thinking came at the expense of rehearsal for the working memory task. Altogether, this demonstrates that processing self-referential information can instigate self-generated thought. Goal competition between self-generated thought and the task at hand provides a likely explanation for how self-generated thought arises.

In Chapter 3, we examined whether embedding participant's current concerns in a task could trigger self-generated thinking. In particular, we were interested in whether embedding personal current concerns could instigate “sticky” thoughts – thoughts that are difficult to disengage from. Research on whether personal current concerns can instigate self-generated thought and sticky thought has not been conclusive (see McVay & Kane, 2013; van Vugt & Broers, 2016). Our study showed that personal concerns lead to higher reports of self-generated thought in comparison to concerns from others or no concerns. Therefore, our research shows that processing of information related to current concerns can result in increases in self-generated thought.

However, processing concerns did not influence the subjective stickiness of thought—how difficult it was to let go of it. A possible explanation for this finding is that goal-directed thoughts may not be experienced as difficult to disengage from, at least, not more than task-focused thoughts. According to the dynamic framework of spontaneous thought (Christoff et al., 2016), sticky thought may be more characterized by automatic constraints arising from the emotional significance of the experienced thought. Whereas goal-directed thoughts would be more deliberately constrained. Future research is needed to determine whether processing of emotionally salient information can instigate sticky thoughts.

The theoretical framework makes two predictions: first, self-generated thought can be manipulated, and second, that self-generated thought can only persist when the resources required for its maintenance remain available. Cognitive resources are assumed to be exclusive to one task at a given time (Salvucci & Taatgen, 2010), therefore, the train of thought will be interrupted and the associated mental model will be bumped out of memory as soon as a task wants to make use of the resources. Existing research on mind wandering has already demonstrated that working memory demand from an ongoing task limits the likelihood of mind wandering occurring (Levinson et al., 2012; Smallwood,

Brown, et al., 2011; Smallwood et al., 2009).

In Chapter 5, we tested the second prediction by examining the influence of the ongoing task (and hence available resources) on the integrity of self-generated thought more thoroughly. In two experiments we manipulated the working memory demands from the ongoing task, and contrasted working memory load to manipulations only targeting visual and declarative memory resources, to investigate how differential demands from the ongoing task influence whether participants prepared for an upcoming task and the extent to which this planning effort was useful. Our results showed that the probability of engaging in advance planning and its productiveness decreased when the ongoing task was more difficult. In other words, complexity alone can influence self-generated thought regardless of the kind of cognitive resource that is being manipulated. However, the impact of working memory load was much more pronounced. Therefore, our research supports that working memory plays a key role in planning ahead, and possibly self-generated thought more generally. Furthermore, we provide support for the theory's prediction that self-generated thought can only effectively exist when its required resources are not occupied by another task.

6.2 Following self-generated thought in task performance and eye tracking measures

Investigating self-generated thought is challenging because it is largely a spontaneous process with few externally observable signs (Smallwood & Schooler, 2015). To obtain insight in self-generated thought, one must therefore (1) know when it occurs during a task and (2) have proper indicators to gain insight in the process. The research in this thesis used subjective thought probes, objective task performance, and eye tracking. The different measures were triangulated to obtain an account of self-generated thought that is not tied to a single (kind of) measure.

In line with existing literature, the research in this thesis showed that probe-caught self-generated thoughts result in performance costs to an ongoing task (Cheyne et al., 2009; Franklin et al., 2011; McVay & Kane, 2009; Smallwood, McSpadden, & Schooler, 2008; van Vugt & Broers, 2016). In Chapter 2, we found that participants were able to recall fewer items on a complex working memory task trial when they reported mental elaboration or mind wandering on that trial. Chapter 3 demonstrated an increase in commission errors to infrequent target stimuli in a sustained attention to response task (SART). Together, these findings support that self-generated thought involves a failure to maintain attention to the task leading to performance loss (McVay & Kane, 2010).

Although the research in Chapter 2 and 3 demonstrated that being engaged in self-

generated thought was associated with clear costs to performance, the research in Chapter 5 showed – interestingly – that it did not. We did not find an increase in errors on the primary task when participants planned ahead for the secondary task. This was true for both experiments and for all difficulty levels of the primary task. We did find that participants were a bit slower to respond on the primary task, but the magnitude of this slowing down was negligible (i.e., 29 ms slower on a ~700ms response latency). That the cost of planning ahead was small-to-nothing is interesting, because it suggests that we are reasonably capable in deciding when it is safe to engage in self-generated thought in given circumstances. That we can modulate self-generated thinking in response to current task demands has been shown before (see Seli, Carriere, et al., 2018; Thomson et al., 2013). Seli and colleagues (2018) performed an experiment in which participants were shown an analog clock. The participants were instructed to push a button every time the clock's hand pointed at 12 o'clock corresponding to a 20 second interval. The researchers found that the mind wandering rates were maximized around halfway the interval (i.e., 6 o'clock), but became smaller when the moment to respond was approaching (i.e., 12 o'clock). In other words, participants engaged in mind wandering more often when they knew that there was room for slacking, but refrained from wandering off when the moment to respond was close. Similar to our study, the propensity of mind wandering was not found to be associated with performance on the task. A difference is that target events were completely predictable in the study of Seli and colleagues, whereas target events were unpredictable in our study. Therefore, we may be able to modulate self-generated thought in response to predictable task demands, as well as in response to situations where it is less certain when response will be required.

Another class of findings in this thesis is derived from eye tracking. The research in Chapter 3 supports existing findings showing that self-generated thought results in a smaller task-evoked response in pupil size (TERPs) compared to on-task thought (Jubera-García et al., 2019; Smallwood, Brown, et al., 2011; Unsworth & Robison, 2016). We similarly found that when participants reported sticky thought, the TERPs were smaller compared to when they reported non-sticky and neutral thinking. The magnitude of TERPs has been associated with the amount of attention, cognitive control, and cognitive processing required by the task (Gilzenrat et al., 2010; Unsworth & Robison, 2014; Wierda et al., 2012). Therefore, the pupil size findings confirm the earlier conclusion that was based on task performance: self-generated thought and particularly a sticky mode of thinking are associated with reduced attention towards the task at hand.

Pupil size can be measured as a transient increase in response to task events, but also in the periods between task events. The average pupil size during such “task-free” periods is called baseline pupil size. Baseline pupil size has been associated with self-generated

thought as well (Unsworth & Robison, 2016), but its magnitude has also been found to correlate with arousal states (Murphy et al., 2014, 2011). So far, many studies have examined the relationship between baseline pupil size and self-generated thought but no consensus has been reached about the direction of the relationship (Grandchamp et al., 2014; Jubera-García et al., 2019; Smallwood, Brown, et al., 2011; Unsworth & Robison, 2016). This thesis adds to this body of work by showing no consistent relationship between self-generated thought and baseline pupil size. Chapter 2 showed that mind wandering was associated with a smaller baseline pupil size, suggesting that mind wandering is associated with a lower state of arousal compared to on-task thought in a demanding task. However, no difference in baseline pupil size was found between mental elaboration about the self-referential material and on-task thinking. In Chapter 3, no direct relationship between self-generated thought (and sticky thought) and baseline pupil size was found in a SART. Instead, the relationship was fully mediated by time-on-task influences. In line with other SART studies, baseline pupil size decreased over the course of the experiment. This decrease in baseline pupil size was observed for on-task thought as well as self-generated thought and other distracted attentional states. At the same time, the frequency of on-task reports was lower towards the end of the task while reports of self-generated thought were higher at that time. Therefore, one might conclude that self-generated thinking is associated with a smaller baseline pupil size, but this conclusion is incorrect. The baseline pupil size was not systematically smaller during self-generated thought compared to on-task thinking at any point during the experiment. Based on the findings from both chapters I conclude that self-generated thought may not have a specific correlate in baseline pupil size, and consequently, no direct relationship with arousal. This conclusion is supported by very recent work showing that a classifier trained on EEG data obtained during a period of low vigilance in a visual search task was not able to predict self-reported mind wandering at above chance level. However, a classifier trained on subjective self-reports of mental state was able to predict mind wandering in the task on the basis of EEG (Jin, Borst, & van Vugt, 2020). This recent study demonstrates that the neural correlate of mind wandering is different from that of a low vigilance/low arousal state.

6.3 Self-generated thought as an adaptive process

The previous sections largely focused on the maladaptive consequences of engaging in self-generated thought for our ongoing activities and what this tells us about the nature of its underlying processes. However, our understanding of self-generated thought would remain limited if we would only investigate it through its negative effects on ongoing tasks. Many studies to this date have indicated that engaging in self-generated thought can have an adaptive function (Baird et al., 2012; Medea et al., 2016; Mooneyham & Schooler, 2013; Smallwood et al., 2013; Steindorf & Rummel, 2017). Self-generated

thought may provide an opportunity to consolidate recent experiences, for creative problem solving, for imagining future events, and possibly much more.

Although it has been demonstrated that engaging in self-generated thought can be adaptive, we do not fully understand yet how and why that is. Existing research has been mostly concerned with classifying the kind of thinking that occurs during a task and studying the consequences of such thinking to the task at hand. Therefore, we have relatively little understanding of the actual content of self-generated thought episodes, how content is selected for self-generated thought, and what processes underlying self-generated thought are responsible for the observed adaptive consequences (Berntsen, 2019).

In Chapter 4, we aimed to answer the question of how and why self-generated thought can have an adaptive function by reviewing several, hitherto separate, literatures. We discussed that the adaptive function of self-generated thought is rooted in the ability to simulate experiences. Self-generated thought is not a process of mere rehearsal of individual memories. Extending our knowledge of episodic future thinking, self-generated thought is a process of flexibly combining and re-combining memories to simulate (novel) experiences (i.e., constructive simulation hypothesis; see e.g., Addis & Schacter, 2008; Hassabis & Maguire, 2007; Schacter & Addis, 2009; Schacter, Addis, & Szpunar, 2017; Schacter, Norman, & Koutstaal, 1998). Which memories are selected during such simulations is not random. We suggest that the selection of memories during self-generated thought is prioritized by its future utility. Research has shown that the likelihood of retrieving a particular fact from memory is associated with its past relevance (determined by recency and frequency of exposure) and its relationship with the current context, such as our current (unfinished) goals (Anderson, 2007; Anderson & Schooler, 1991; Klinger, 2013; Schooler & Anderson, 1997; Steindorf & Rummel, 2017). By prioritizing the retrieval of memories with future utility, we selectively strengthen memory structures with future relevance, essentially preparing our memory for the future.

Apart from strengthening memory structures with future relevance, another important consequence of simulating experiences during self-generated thought is that it may allow for construction of (potentially) novel memory structures. Flexibly combining and re-combining memories during self-generated thought may allow for novel connections to be drawn between memories, resulting in updated or new memories. This process of creating novel bindings between memories has been proposed to depend on working memory (Oberauer, 2005; Oberauer et al., 2013). One particular activity during self-generated thought that may be supported by the ability to construct new memory

structures is planning (e.g., Cole, Bagic, Kass, & Schneider, 2010; Cole, Patrick, Meiran, & Braver, 2018; Oberauer, Souza, Druey, & Gade, 2013; Pereg & Meiran, 2019). Planning can be defined as a process in which we identify the necessary steps towards achieving a future goals (Spreng et al., 2015). Depending on the goal, planning would involve identifying and sequencing which tasks, operations, and/or factual concepts are needed. In Chapter 5, we demonstrated that self-generated thought can support planning ahead. Participants performed a sequential dual-task paradigm in which they could (voluntarily) prepare for a second task while performing a first task. The results showed that participants could productively plan ahead with little cost to the first task. However, the demands from the ongoing first task limited to what extent the participants could productively plan ahead. In line with the idea that constructing new memory structures depends on the availability of working memory, we found that working memory demand from the first task reduced the benefits of planning ahead.

All in all, we propose that self-generated thought can have an adaptive function because of two central mechanisms or principles. First of all, the selection of memories for self-generated thought is prioritized for future utility. These retrieved memories can subsequently be used to strengthen, update or create novel memories structures with potential future relevance.

6.4 Conclusion

In this thesis we investigated the kinds of thought that comes up while already being engaged in a task. Such self-generated thoughts are not a nuisance, but occupy a large portion of our waking lives. A process that consumes much of our conscious experience must play a key role in our daily functioning. This thesis provided new insights on the mechanisms that drive self-generated thinking and its consequences for ongoing activities, both positive and negative.

In short, we learned that self-generated thought is a process that competes for limited cognitive resources with the ongoing task (and potentially other tasks). The availability of working memory and declarative memory at a particular moment creates the opportunity for self-generated thought to arise. Conversely, engaging in self-generated thought limits working memory and declarative memory use for the ongoing activity, potentially resulting in costs. The likelihood for self-generated thought to arise, and therefore to win the competition, can be increased by processing self-referential material and by being exposed to personal concerns.

In essence, self-generated thought is a process of memory retrieval and processing these memories for retention or elaboration. Which memories are retrieved during episodes

of self-generated thought is not random. The selection of memories for self-generated thought is prioritized for future relevance determined a history of past relevance and association with current (unfinished) goals. Subsequently, the retrieved memories can be used to strengthen, update or construct novel memories structures with potential future utility. These two mechanisms, the prioritization of memories for self-generated thought and the ability to strengthen and construct memory structures during self-generated thought, may allow for adaptive functioning of self-generated thought.

6.5 Practical implications

The research in this thesis demonstrates that self-generated thought is a very useful activity. Engaging in self-generated thought provides an opportunity to prepare for upcoming activities. Other research has suggested that engaging in self-generated thought may facilitate creative problem solving, prospective remembering and sensemaking (Baird et al., 2012; Christoff et al., 2011; Leszczynski et al., 2017; Szarras & Niedźwieńska, 2011; Tan et al., 2015). Therefore, despite the negative connotations that self-generated thinking has, it is well-advised to engage in self-generated thought from time to time.

Whether engaging in self-generated thought will be adaptive is dependent on present task conditions. Our research suggests that engaging in self-generated thought can only yield adaptive consequences when a present task does not concurrently demand declarative memory and working memory. Examples of such moments are when we engage in routine tasks, such as cycling our daily route to work or school, washing the dishes, and vacuum cleaning the house. These tasks only require procedural, motor and peripheral (visual or auditory) resources. However, a large portion of our day we engage in activities that are not routine. At school or at work, we are often required to encode information, memorize information, and solve problems with that information. Engaging in self-generated thought during such moments will likely result in performance loss on the ongoing activity, or, limit the possible benefits gained from adaptive self-generated thought.

When engaging in non-routine activities, we recommend having short periodic breaks to allow self-generated thought to occur in moments of rest. A popular and easy-to-use method to incorporate short breaks in activities is the Pomodoro technique (Cirillo, 2018). The Pomodoro technique asks you to break down an activity in four sprints of 25 minutes (more or less sprints is possible). Each sprint is interleaved with a break of 5 minutes. After four sprints a longer break 20 minutes is suggested. The research in this thesis provides some support for this technique. In our studies, we frequently observed that reports of self-generated thought increased over time. Experiments in

this thesis were usually around one hour, but frequent reports of self-generated thinking were already observed at around 20 minutes in the experiment. Therefore, sprints of 25 minutes seem fitting. The five-minute breaks are short, but do allow some time to let our minds wander without breaking too much the flow of our activity—at least if we do not immediately occupy it with other content such as social media. The longer break allows for more elaborated self-generated thought.

The effectiveness of periodic short breaks is highly dependent on our ability to defer self-generated thought to these short breaks. This thesis did not explicitly examine how this can be accomplished or how we can get better at doing this. Nonetheless, some things can be said about it. First of all, our research shows that we are already reasonably good at strategically moderating self-generated thought with respect to ongoing task demands. However, we might get better at deferring self-generated thought when we achieve more control over the goal competition process. Goal competition is an automatic process, but the outcome of the competition can be manipulated in favor of a present task when its associated goals are kept highly active. Possibly, one promising way to achieve better goal maintenance is to practice mindfulness, specifically the focused-attention (FA) type (Mrazek, Smallwood, & Schooler, 2012). Currently, there is some initial evidence that mindfulness can reduce self-generated thought (Bennike, Wieghorst, & Kirk, 2017; Mrazek, Franklin, Phillips, Baird, & Schooler, 2013; Vanucci & Agnoli, 2019; Xu, Purdon, Seli, & Smilek, 2017). The practice of mindfulness is straightforward: 1) you bring your attention to an object of focus such as your breath, 2) while focusing on the object you monitor your attention without judgement, 3) you bring your attention back to the object of focus when your attention has slipped. According to Wallace (1999), the practice of mindfulness helps to refine our ability to remember to sustain attention to an object and to remember to bring our attention back to the object of focus when attention has wandered off. In other words, practicing mindfulness promotes goal checking and rehearsal of currently active goals (Moye & Van Vugt, 2021), all of which contribute to goal maintenance, and in turn – defer self-generated thought.

NEDERLANSE SAMENVATTING

Misschien heb je er nog nooit bij stil gestaan: gedurende de dag heb je – vrijwel – constant gedachten. Ons brein lijkt nooit te stoppen met denken. We zijn met onze gedachten bij de dingen die we doen op ons werk, op school of thuis. Maar ook als we even ‘niets’ doen is ons hoofd niet stil. We dagdromen over een mooie vakantie die voor de deur staat, we plannen alvast vooruit voor zaken die binnenkort belangrijk zijn of denken terug aan de gebeurtenissen van afgelopen weekend. Deze gedachten die niet gericht zijn op wat we op dit moment aan het doen zijn en wat er nu in onze omgeving gebeurt noemen we *zelf-gegenereerde gedachten*¹.

Zelf-gegenereerde gedachten zijn geen toevallige bijkomstigheid. Wetenschappers denken zelfs dat we tot de helft van onze dag spenderen aan deze gedachten. Een proces dat zoveel tijd in beslag neemt heeft noodzakelijk gevolgen voor het dagelijks functioneren. Onderzoeken tonen aan dat dagdromen op school samenhangt met slechtere leerprestaties. In het verkeer is dagdromen een belangrijke factor bij het ontstaan van ongelukken. De vorm en inhoud van zelf-gegenereerde gedachten heeft effect op ons functioneren en welzijn. Met name zelf-gegenereerde gedachten met een rigide en inflexibel karakter, zoals piekeren, hangen samen met verslechtering van stemming en vormen een risicofactor voor het ontstaan van depressie en andere gemoedstoornissen. Toch is het hebben van zelf-gegenereerde gedachten niet altijd slecht. Integendeel! Onderzoek toont aan dat dagdromen met een positieve inhoud ons beter kan later voelen. Zelf-gegenereerde gedachten bieden tevens ook de mogelijkheid om te reflecteren, om voorstellingen te maken over toekomstige situaties, vooruit te plannen en creatieve oplossingen te bedenken. Iedereen heeft waarschijnlijk wel een keer heeft meegemaakt dat de oplossing voor een lastig probleem opkwam tijdens het dagdromen.

Het hebben van zelf-gegenereerde gedachten is dus veelvoorkomend fenomeen. Daarnaast heeft het zowel positieve als negatieve gevolgen. Omdat zelf-gegenereerde gedachten invloed hebben op ons dagelijks functioneren, is het belangrijk om goed te begrijpen hoe deze gedachten werken. In dit proefschrift was daarom de doelstelling om te onderzoeken hoe zelf-gegenereerde gedachten werken. Daarnaast wilden we onderzoeken waarom en wanneer zelf-gegenereerd denken positieve of negatieve

¹ In de volksmond wordt vaak de term dagdromen gebruikt. Alhoewel dagdromen een onderdeel is van zelf-gegenereerde gedachten, is de term ‘dagdromen’ te beperkt. Onder dagdromen wordt veelal een spontaan en associatief denkproces verstaan. Hoewel zelf-gegenereerde gedachten dit karakter kunnen aannemen, kunnen zelf-gegenereerde gedachten ook actief gestuurd en doelgericht zijn. ‘Zelf-gegenereerde gedachten’ is dus een alomvattende term voor de gedachten die niet op de omgeving gericht zijn. Dagdromen is een specifieke uitingvorm van zelf-gegenereerde gedachten.

gevolgen heeft.

In deze samenvatting wordt eerst toegelicht hoe zelf-gegenereerde gedachten in dit proefschrift zijn onderzocht. Daaropvolgend worden de resultaten van dit onderzoek beschreven. De samenvatting sluit af met een conclusie op basis van de resultaten.

Onderzoeksmethode

Om psychologische processen, zoals zelf-gegenereerde gedachten, goed te kunnen onderzoeken zijn er in ieder geval twee zaken belangrijk. Ten eerste is het van belang om een zekere controle te hebben over het proces dat wordt onderzocht. Met controle wordt bedoeld dat beïnvloed kan worden wanneer het proces in meer of mindere mate plaatsvindt. Bijvoorbeeld, bij het onderzoeken van geheugen kunnen in een experiment getallen gegeven worden om te onthouden. De mate waarin geheugen in het experiment nodig is, kan worden beïnvloed door het aantal getallen dat onthouden moet worden te variëren. Ten tweede is het belangrijk om het proces goed te kunnen meten. Bijvoorbeeld, in het eerdere voorbeeld over geheugen is het aantal correct gerapporteerde getallen een maat voor de sterkte van het geheugen. Bij het onderzoeken van zelf-gegenereerde gedachten zijn er bij beide zaken uitdagingen. Het is lastig om controle over zelf-gegenereerde gedachten te hebben, omdat het een veelal een spontaan proces is. Een persoon kan niet zomaar gevraagd worden om te dagdromen. Daarnaast is het lastig om objectief van buiten te zien of iemand zelf-gegenereerde gedachten heeft. Er zijn maar weinig observeerbare kenmerken waaraan afgeleid kan worden of iemand op een bepaald moment zelf-gegenereerde gedachten heeft.

Met de bovenstaande uitdagingen in het achterhoofd zijn in ieder geval twee dingen nodig om zelf-gegenereerde gedachten te kunnen onderzoeken. Allereerst is een taak nodig waarin mogelijkheid is om zelf-gegenereerde gedachten te laten plaatsvinden. Wanneer zelf-gegenereerde gedachten plaatsvinden in een taak, kan maar beperkt worden gecontroleerd, maar op zijn minst kan de taak wel zo ingericht worden dat er ruimte voor zelf-gegenereerd denken wordt geboden. In de praktijk betekent dit dat een taak (korte) periodieke onderbrekingen moet bevatten waarin tijd en ruimte is om zelf-gegenereerde gedachten te laten ontstaan. Een voorbeeld is te zien in Figuur 1 (bovenste paneel). In Figuur 6.1 wordt een klassiek experiment afgebeeld die ook in hoofdstuk 3 van dit proefschrift is gebruikt. In dit klassieke experiment voeren deelnemers een eenvoudige taak uit op een computer. De deelnemers krijgen de taak om zodra ze een woord zien op het scherm dat in kleine letters is geschreven zo snel mogelijk op spatie te drukken. Echter, wanneer een woord in hoofdletters is geschreven, dan moeten ze niet drukken. Tussen de presentatie van de woorden vindt er een korte taakonderbreking plaats. Deze taakonderbreking vraagt van de deelnemer om te blijven focussen, maar

**ZELF-GENEREERDE GEDACHTEN
ONDERZOEKEN: HOE DOE JE DAT?**

Experimentele taak:

Taakonderbrekingen geven ruimte voor zelf-
gegenereerde gedachten

Zelf-rapportage:

Opvangen van gedachten met periodieke thought probes

Thought probe

Triangulatie:

Inzicht in zelf-gegenereerde gedachten door
combineren van zelf-rapportage,
taakprestaties en eye tracking

Figuur 6.1. Overzicht van hoe dagdromen onderzocht kan worden in een experimentele setting. Het bovenste paneel toont een voorbeeld van hoe taakonderbrekingen ingezet kunnen worden om ruimte te geven aan zelf-gegenereerde gedachten. Het middelste paneel toont een voorbeeld van hoe zelf-gegenereerde gedachten gemeten kunnen worden aan de hand van zelfrapportage. Het onderste paneel toont hoe inzicht in zelf-gegenereerde gedachten verkregen kan worden door triangulatie.

biedt tegelijkertijd ook ruimte om af te dwalen².

Ten tweede zijn meetmethoden nodig om zelf-gegenereerde gedachten op te vangen wanneer het plaatsvindt gedurende een taak. In twee van de drie experimenten in dit proefschrift hebben we dit gedaan aan de hand van zelfrapportage. Met deze methode krijgen deelnemers aan een experiment op willekeurige momenten in het experiment een vraag voorgelegd, ook wel een *thought probe* genoemd (zie Figuur 6.1, middelste paneel). De vraag luidt: ‘waar was u zonet met uw gedachten?’. De deelnemers kunnen dan antwoorden met verschillende opties, waaronder in ieder geval een optie voor gedachten die gericht zijn op de uitvoering van het experiment of gericht zijn op andere zaken en dus zelf-gegenereerd zijn.

Met periodieke zelf-rapportage kan worden gemeten wanneer de deelnemers in een experiment afgedwaald zijn of niet. Echter, met deze kennis is er nog geen inzicht in hoe zelf-gegenereerde gedachten werken en wat de gevolgen van zelf-gegenereerde gedachten zijn voor een huidige (ofwel toekomstige) activiteit. Daarvoor moet zelf-rapportage gecombineerd worden met andere meetmethodes. Het combineren van meetmethodes heet *triangulatie* (zie Figuur 6.1, onderste paneel). In dit proefschrift hebben we zelf-rapportage gecombineerd met metingen van *taakprestaties* en met *eye tracking*. Voorbeelden van metingen van taakprestaties zijn het aantal goede/foute antwoorden dat de deelnemer geeft en de reactietijd bij het antwoorden. Vanuit deze metingen kan worden afgeleid hoe zelf-gegenereerde gedachten werken en wat de gevolgen zijn. Wanneer meer foute antwoorden en een veranderde reactietijd wordt geobserveerd, kort voordat de deelnemer op de *thought probe* antwoordt dat hij of zij was afgedwaald, dan is dat een indicatie dat zelf-gegenereerde gedachten negatieve gevolgen heeft in die situatie. Tegelijkertijd geeft dat ook inzicht in hoe zelf-gegenereerde gedachten werken. Wanneer zelf-gegenereerde gedachten de taakprestaties negatief beïnvloeden, dan kan daaruit worden afgeleid dat er overlap is tussen wat nodig is om de taak goed te doen en wat er nodig is om te kunnen dagdromen.

Eye tracking is een meetmethode die het mogelijk maakt om continu vast te leggen waar mensen kijken op een scherm (d.w.z. oogbewegingen) en hoe groot hun pupil op dit moment is (d.w.z. pupilgrootte). In dit proefschrift hebben we zowel oogbewegingen als pupilgrootte gemeten voor verschillende doeleinden. Oogbewegingen zijn nuttig omdat de locatie van iemands blik een redelijke indicatie geeft van waar de aandacht van de persoon op gericht is en wat zijn of haar doelen zijn. Zijn de reeksen van blikken die we observeren een indicatie of de deelnemer nog betrokken is bij de taak die we hebben

² Het is wellicht goed te vermelden dat de deelnemer het experiment uitvoert afgezonderd van de onderzoeker.

gegeven of niet? Als we bijvoorbeeld iemand naar verschillende producten op het schap van een supermarkt ziet staren, is het redelijk om te concluderen dat deze persoon op zoek is naar een bepaald product. Wanneer verwacht gedrag van oogbewegingen, gegeven een taak, niet wordt waargenomen, kan dit erop wijzen dat een persoon bezig is met zelf-gegenereerde gedachten. Pupilmagnitude hebben we gemeten omdat aandacht en gedachtenprocessen invloed hebben op hoe groot onze pupil is op een bepaald moment. Wanneer informatie uit onze omgeving, of in een experimentele taak, wordt verwerkt, dan zorgt dat voor korte vergroting in de pupil. De mate van deze vergroting is afhankelijk van hoeveel aandacht de persoon stopt in dit verwerkingsproces. De grootte van iemands pupil over een periode is ook informatief. Een gemiddeld kleine pupil grootte is een indicator dat iemand in een onoplettende staat is. Een gemiddeld grote pupilmagnitude is een indicator voor een overactieve afleidbare staat. De gemiddelde pupilmagnitude kan worden gebruikt om te onderzoeken of zelf-gegenereerde gedachten vooral ontstaan in een onoplettende of overactieve staat.

Resultaten

Zelf-gegenereerde gedachten als een competitief proces

Zoals eerder vermeld in deze samenvatting is de doelstelling van dit proefschrift om meer inzicht te geven in hoe zelf-gegenereerde gedachten werken. Daarnaast wil dit proefschrift meer inzicht geven in de consequenties van zelf-gegenereerde gedachten voor huidige en toekomstige activiteiten. In de volgende paragrafen worden de resultaten van dit proefschrift samengevat.

Het onderzoek is gestart met het opstellen van een theoretisch kader, waarmee precieze voorspellingen gedaan konden worden over hoe zelf-gegenereerde gedachten kunnen ontstaan en voortduren in de context van een huidige activiteit. Dit theoretische kader bevat twee belangrijke proposities over zelf-gegenereerde gedachten. Allereerst zijn zelf-gegenereerde gedachten een gewoon denkproces dat dezelfde (of op zijn minst zeer vergelijkbare) cognitieve processen gebruikt als elk ander denkproces. Ten tweede stelt het theoretisch kader dat zelf-gegenereerde gedachten kunnen plaatsvinden parallel aan andere denkprocessen, zonder dat dit consequenties heeft voor de processen. Het kader voorspelt dat zelf-gegenereerde gedachten kunnen ontstaan en voortduren wanneer de vereiste middelen beschikbaar zijn voor gebruik. In dit proefschrift gingen we ervan uit dat zelf-gegenereerde gedachten aanspraak doen op ons langetermijngeheugen voor het ophalen van herinneringen en op werkgeheugen om de opgehaalde herinneringen tijdelijk in op te slaan. Of zelf-gegenereerde gedachten daadwerkelijk plaatsvinden wanneer langetermijngeheugen en werkgeheugen beschikbaar zijn voor gebruik, is afhankelijk van een aanvullend competitieproces. Deze competitie vindt plaats tussen zelf-gegenereerde gedachten en de taak die voor handen is die dezelfde middelen wil

gebruiken. Welk proces de concurrentie wint, hangt af van de relatieve prioriteit van deze processen op een bepaald moment. Kortom, wanneer middelen voor zelf-gegenereerde gedachten beschikbaar zijn, zullen processen van verschillende taken en doelen hier om strijden. De relatieve prioriteit van deze processen zal bepalen waar vervolgens over wordt nagedacht.

De prioriteit van mentale processen wordt mede bepaald door de interne en externe omgeving. De interne omgeving omvat onze huidige doelen, opgehaalde herinneringen en de inhoud van ons werkgeheugen. De externe omgeving omvat wat we waarnemen met onze zintuigen, met name wat we zien en horen. Het feit dat de omgeving invloed kan uitoefenen op de prioriteit van mentale processen betekent dat zelf-gegenereerde gedachten mogelijk ook waarschijnlijker gemaakt kunnen worden door afleidende informatie aan te bieden in de interne en externe omgeving. In **hoofdstuk 2** hebben we deze voorspelling getoetst door te onderzoeken of het verwerken van informatie refererend aan onszelf (bijv. ‘Ben ik aardig?’, ‘Ben ik achterdochtig?’) het competitieproces kan beïnvloeden ten bate van zelf-gegenereerde gedachten. In dit hoofdstuk vond het verwerken van de zelf-refererende informatie plaats in de context van een werkgeheugentaak, waarin een serie van ruimtelijke locaties onthouden moesten worden. Deelnemers aan het onderzoek zagen in een opgave steeds eerst een locatie om te houden, gevolgd door een serie van zelf-refererende informatie en een taakonderbreking. Tijdens deze taakonderbreking kregen deelnemers de kans om de locatie die ze kort geleden gezien hadden in te studeren (herhalen in het hoofd) of af te dwalen in zelf-gegenereerde gedachten (zie ook Figuur 6.1). We voorspelden dat het verwerken van zelf-refererende informatie de kans op het ontstaan van zelf-gegenereerde gedachten kon vergroten tijdens de taakonderbrekingen, ten koste van het instuderen van de locaties. Onderzoek heeft aangetoond dat mensen de neiging hebben om door te blijven denken over zelf-refererende informatie. Bovendien kan het verwerken van zelf-refererende informatie mogelijk zelf-gegenereerde gedachten opwekken door associaties met persoonlijke doelen en aspiraties. De resultaten van ons onderzoek lieten zien dat deelnemers inderdaad meer zelf-gegenereerde gedachten hadden na het verwerken van zelf-refererende informatie. De toename van zelf-gegenereerde gedachten ging ten koste van het instuderen van de locaties, en daarmee de prestaties op de geheugentaak. Gezamenlijk toont dit aan dat een competitieproces een plausibele verklaring is voor wanneer en waarom zelf-gegenereerde gedachten ontstaan.

In **hoofdstuk 3** hebben we onderzocht of het verwerken van informatie gerelateerd aan persoonlijke acute zorgen zelf-gegenereerd denken kan triggeren. Voorbeelden van persoonlijke acute zorgen zijn het vinden van een nieuwe baan, het afmaken van een

afstudeeronderzoek en het plannen van een vakantie. Hierbij is informatie gerelateerd aan de zorgen. Bijvoorbeeld een vacaturetekst in de krant, een afspraak met een begeleider in je agenda en het zien van een reisbureau. In hoofdstuk 3 waren we met name geïnteresseerd in de vraag of persoonlijke acute zorgen kunnen leiden tot sticky gedachten. Sticky gedachten zijn gedachten waarvan je ervaart dat ze lastig los te laten zijn. We hebben dit onderzocht door de persoonlijke acute zorgen van deelnemers te verwerken in een taak. De taak die de deelnemers uitvoerden was heel vergelijkbaar met de taak die we eerder hebben laten zien in Figuur 6.1. In deze taak moesten deelnemers op spatie drukken wanneer een woord in kleine letters was geschreven, en niet wanneer een woord in hoofdletters was geschreven. Persoonlijke acute zorgen van de deelnemers werden verzameld aan de hand van een vragenlijst. De persoonlijke zorgen die als meest belangrijk werden aangegeven door de deelnemers, hadden we vertaald naar een drietal woorden en vervolgens als woorden in de taak geplaatst. Bijvoorbeeld de zorg voor het vinden van een nieuwe baan werd in de taak gebruikt als ‘zoeken – nieuw – carrière’. Kort na de presentatie van de zorg werd een thought probe gepresenteerd. Om te kunnen onderzoeken of de zorgen leiden tot meer zelf-gegenereerd en meer sticky denken werden er ook zorgen van andere deelnemers in de taak geplaatst. Ons onderzoek toonde aan dat deelnemers vaker zelf-gegenereerde gedachten rapporteerden kort nadat persoonlijke zorgen waren getoond, in vergelijking met wanneer de zorgen van anderen waren getoond of helemaal geen zorgen. Daarom laat ons onderzoek zien dat het verwerken van informatie met betrekking tot actuele persoonlijke zorgen kan resulteren in een toename van zelf-gegenereerde gedachten. Het verwerken van persoonlijke zorgen had echter geen invloed op hoe sticky de gedachten waren. Een mogelijke verklaring hiervoor is dat gedachten gericht op persoonlijke zorgen niet worden ervaren als lastiger om los te laten dan gedachten gericht op het uitvoeren van een taak.

Het theoretisch kader doet twee voorspellingen over zelf-gegenereerde gedachten. Ten eerste dat informatie in de interne en externe omgeving het ontstaan van zelf-gegenereerde gedachten kan beïnvloeden. Ten tweede dat zelf-gegenereerde gedachten alleen kunnen voortduren wanneer de noodzakelijke middelen voor het proces beschikbaar blijven. Dezelfde (cognitieve) middelen kunnen namelijk door maar één taak of proces tegelijk worden gebruikt. Ter illustratie, we kunnen niet naar onze telefoon kijken en het verkeer tegelijk in de gaten houden (visuele middelen). We kunnen ook niet tegelijk een som uitrekenen en een tel bijhouden (langetermijngeheugen en werkgeheugenmiddelen). In **hoofdstuk 5** hebben we deze tweede voorspelling onderzocht door te kijken naar de invloed van een huidige taak (en dus de overblijvende beschikbare middelen) op de integriteit van zelf-gegenereerde gedachten. In twee experiment beïnvloedden we de hoeveelheid werkgeheugenbelasting vanuit de huidige taak en vergeleken we

variaties in werkgeheugenbelasting met situaties waarin alleen visuele middelen of langtermijngeheugen werd belast. Dit om te onderzoeken of variatie in belasting vanuit de huidige taak invloed heeft op ‘of’ en ‘wanneer’ we plannen voor een aankomende taak en hoe productief dit plannen is. De resultaten van dit onderzoek lieten zien dat de waarschijnlijkheid dat deelnemers vooruit gingen plannen voor de aankomende taak en het voordeel dat dit vooruitplannen opleverde voor de aankomende taak achteruit ging met groeiende belasting vanuit de lopende taak. Met andere woorden, belasting vanuit een huidige taak, onafhankelijk van wat voor soort belasting het is, kan al beïnvloeden of en wanneer zelf-gegenereerde gedachten ontstaan en hoeveel voordeel we eruit halen. Echter, ons onderzoek toonde ook aan dat werkgeheugenbelasting de sterkste invloed had. Daarom kunnen we concluderen dat werkgeheugen mogelijk een belangrijke rol speelt in vooruitplannen en mogelijk in zelf-gegenereerde gedachten in het algemeen. Daarnaast ondersteunen de resultaten het theoretisch kader dat voorspelt dat zelf-gegenereerde gedachten alleen kunnen bestaan wanneer de nodige middelen beschikbaar zijn.

Zelf-gegenereerde gedachten in taakprestaties en eye tracking

Wanneer we zelf-gegenereerde gedachten hebben tijdens een huidige activiteit, bijvoorbeeld tijdens het lezen van deze samenvatting, dan is het een logische verwachting dat dit consequenties heeft voor de huidige activiteit. Inderdaad, het onderzoek in dit proefschrift toonde aan dat zelf-gegenereerde gedachten vaak tot prestatieverlies leidde op een huidige activiteit. In **hoofdstuk 2** konden deelnemers minder locaties onthouden in opgaven waarbij ze hadden aangegeven (aan de hand van een thought probe) dat ze zelf-gegenereerde gedachten hadden. Ook in **hoofdstuk 3** maakten de deelnemers meer fouten op de taak. Dat wil zeggen, ze drukten per ongeluk op spatie terwijl dat niet moest wanneer ze zelf-gegenereerde gedachten hadden. Al met al toont dit aan dat zelf-gegenereerde gedachten gepaard gaan met een verlies van aandacht voor de huidige taak en dat dit leidt tot prestatieverlies.

Alhoewel het onderzoek in hoofdstuk 2 en 3 duidelijk lieten zien dat zelf-gegenereerd denken negatieve gevolgen voor de huidige taak heeft, bleek dit in **hoofdstuk 5** niet het geval. Deelnemers maakten niet meer fouten op de huidige taak wanneer ze tegelijk ook aan het voorbereiden waren op de aankomende taak. Hierbij maakte het bovendien niet uit wat de belasting vanuit de huidige taak was. Wanneer we er ook weer bij betrekken dat deelnemers minder snel geneigd waren om voor te bereiden wanneer de huidige taak belastend was, dan kunnen we concluderen dat wij mogelijk redelijk goed zijn in het inschatten wanneer het veilig is om zelf-gegenereerde gedachten te hebben en wanneer niet.

Naast taakprestaties is ook eye tracking in dit proefschrift gebruikt om inzicht te krijgen in zelf-gegenereerd denken. **Hoofdstuk 3** liet zien dat de vergroting in de pupil als reactie op gebeurtenissen in de taak kleiner was wanneer deelnemers zelf-gegenereerde gedachten of sticky gedachten hadden. Zoals eerder in deze samenvatting vermeld, indexeert de vergroting in de pupil hoeveel aandacht wordt gestopt in het verwerkingsproces. Dus, de resultaten van pupilgrootte sluiten aan bij de eerdere conclusie op basis van taakprestaties, namelijk dat zelf-gegenereerd denken een proces is dat aandacht onttrekt van de huidige taak.

Pupilgrootte kan worden gemeten als reactie op een gebeurtenissen in de taak, maar ook gedurende taakonderbrekingen. De gemiddelde pupilgrootte tijdens taakonderbrekingen indexeert iemands mentale staat: een relatief kleine gemiddelde pupilgrootte is een indicator van een onoplettende en slaperige toestand, een gemiddeld grote pupilgrootte is een indicator van een overactieve afleidbare toestand. Met het meten van de gemiddelde pupilgrootte kan dus worden onderzocht met welke mentale staat zelf-gegenereerd denken samenhangt. Het huidige onderzoek naar de samenhang tussen zelf-gegenereerd denken en gemiddelde pupilgrootte heeft nog geen consensus bereikt. Het onderzoek in dit proefschrift ondersteunt bestaande onderzoeken die hebben aangetoond dat er geen consistente samenhang is tussen gemiddelde pupilgrootte en zelf-gegenereerd denken. Zelf-gegenereerde gedachten ontstaan dus niet met voorkeur in een onoplettende/slaperige staat of een overactieve/afleidbare staat. Het is een mogelijkheid dat het karakter van zelf-gegenereerd denken verschillend is, afhankelijk van de mentale staat waarin iemand verkeerd, echter, dit is in dit proefschrift niet onderzocht.

Zelf-gegenereerde gedachten: een adaptief proces?

De vorige paragrafen focusten met name op de negatieve gevolgen van zelf-gegenereerde gedachten en wat hieruit geleerd kan worden over hoe zelf-gegenereerd denken werkt. Echter, ons begrip van zelf-gegenereerd denken zou erg beperkt blijven als alleen de negatieve kant wordt onderzocht. Meerdere bestaande onderzoeken hebben al reeds aangetoond dat zelf-gegenereerd denken een positieve kant heeft. Zelf-gegenereerd denken geeft de mogelijkheid om recente ervaringen te verwerken, om creatieve oplossingen te bedenken, om toekomstige gebeurtenissen voor te stellen en mogelijk veel meer.

Alhoewel het bekend is dat zelf-gegenereerd denken een positieve kant heeft, is het onduidelijk waarom en hoe. In **hoofdstuk 4** hebben we daarom deze vraag onderzocht. Het onderzoek in hoofdstuk 4 werd niet uitgevoerd aan de hand van een experiment, maar door onderzoeksliteratuur te beschouwen verwant aan zelf-gegenereerd denken.

We beschreven in dit hoofdstuk dat zelf-gegenereerd denken zijn positieve gevolgen ontleend aan de mogelijkheid om ervaringen te simuleren. Zelf-gegenereerd denken is niet alleen maar herinneringen ophalen. Zelf-gegenereerd denken is een proces waarin de opgehaalde herinneringen flexibel gecombineerd kunnen worden om nieuwe ervaringen te simuleren. Welke herinneringen geselecteerd worden tijdens dergelijke simulaties is niet willekeurig. In hoofdstuk 4 suggereren we dat de selectie van herinneringen voor zelf-gegenereerde gedachten geprioriteerd wordt aan de hand van verwachte bruikbaarheid in de toekomst. Bestaand onderzoek laat zien dat de kans op het ophalen van een bepaalde herinnering geassocieerd is aan de bruikbaarheid van de herinnering in het verleden en de relatie met de huidige context (zoals onze huidige persoonlijke doelen). Door herinneringen met bruikbaarheid voor de toekomst te prioriteren, neemt de kans dus toe dat we gedachten hebben die relevant zijn voor onze toekomst. Tegelijkertijd versterkt het ophalen van de herinneringen het geheugen voor deze herinneringen. Dit bereidt in wezen ons geheugen voor op de toekomst.

Afgezien van het versterken van herinneringen die relevant zijn voor de toekomst, is een ander belangrijk gevolg van het simuleren van ervaringen tijdens zelf-gegenereerd denken dat het de constructie van (potentieel) nieuwe geheugenstructuren mogelijk maakt. Door herinneringen flexibel te combineren en opnieuw te combineren tijdens zelf-gegenereerd denken, kunnen nieuwe verbindingen tussen herinneringen worden gemaakt, wat resulteert in bijgewerkte of nieuwe geheugenstructuren. Onderzoeken stellen dat dit proces van het creëren van nieuwe bindingen tussen herinneringen afhankelijk is van het werkgeheugen. Een specifieke activiteit tijdens zelf-gegenereerde gedachten die mogelijk wordt ondersteund door het vermogen om nieuwe geheugenstructuren te construeren, is plannen. Plannen kan worden gedefinieerd als een proces waarin we de noodzakelijke stappen identificeren om toekomstige doelen te bereiken. Afhankelijk van het doel omvat plannen het identificeren en volgen van welke acties gedaan moeten worden. In **hoofdstuk 5** hebben we aangetoond dat zelf-gegenereerde gedachten gebruikt kunnen worden om te plannen. Deelnemers voerden een taak uit waarin ze zich (vrijwillig) konden voorbereiden op een tweede taak, terwijl ze een lopende taak uitvoerden. De resultaten toonden aan dat deelnemers productief vooruit konden plannen met weinig kosten voor de lopende taak. De belasting vanuit de lopende taak beperkte echter wel in hoeverre de deelnemers productief vooruit konden plannen. In overeenstemming met het idee dat het bouwen van nieuwe geheugenstructuren afhankelijk is van de beschikbaarheid van werkgeheugen, ontdekten we dat de vraag naar werkgeheugen vanuit de lopende taak de voordelen van vooruit plannen verminderde.

Al met al concluderen we dat zelf-gegenereerd denken positieve consequenties kan

hebben vanwege twee centrale processen of principes. Allereerst dat herinneringen met nut voor de toekomst worden geprioriteerd tijdens zelf-gegenereerd denken. Ten tweede dat deze opgehaalde herinneringen vervolgens gebruikt kunnen worden om nieuwe geheugenstructuren met mogelijk toekomstig nut te versterken, bij te werken of te creëren.

Conclusie

In dit proefschrift hebben we de gedachten onderzocht die opkomen terwijl we al iets aan het doen zijn. Dergelijke zelf-gegenereerde gedachten zijn geen zeldzaamheid, maar komen veel voor. Een proces dat veel van onze tijd verbruikt speelt waarschijnlijk een sleutelrol in ons dagelijks functioneren. Dit proefschrift heeft nieuwe inzichten opgeleverd over hoe zelf-gegenereerd denken werkt en de gevolgen die het heeft voor lopende en toekomstige activiteiten.

We hebben geleerd dat zelf-gegenereerd denken een proces is dat concurreert om beperkte middelen met een huidige activiteit. De beschikbaarheid van werkgeheugen en langetermijngeheugen op een bepaald moment, creëert de mogelijkheid dat zelf-gegenereerde gedachten kunnen ontstaan. Omgekeerd beperkt zelf-gegenereerd denken het werkgeheugen en het langetermijngeheugengebruik voor een lopende activiteit, wat kan resulteren in prestatieverlies voor de lopende activiteit. De kans dat zelf-gegenereerde gedachten ontstaan, kan worden vergroot door het verwerken van informatie gerelateerd aan ons zelf-concept en onze acute persoonlijke zorgen.

In wezen is zelf-gegenereerd denken een proces van het ophalen van herinneringen. Welke herinneringen worden opgehaald tijdens zelf-gegenereerde gedachten is niet willekeurig. Het selectieproces van herinneringen voor zelf-gegenereerde gedachten prioriteert op toekomstige bruikbaarheid. De toekomstige bruikbaarheid van een herinnering wordt bepaald door geschiedenis van verleden bruikbaarheid en associatie met huidige doelen. Vervolgens kunnen de opgehaalde herinneringen worden gebruikt om geheugenstructuren te versterken, bij te werken of te bouwen met mogelijk toekomstig nut. Deze twee mechanismen, het prioriteren van herinneringen voor zelf-gegenereerde gedachten en het versterken en construeren van geheugenstructuren tijdens zelf-gegenereerd denken, maken mogelijk dat zelf-gegenereerd denken adaptief kan worden gebruikt.

ACKNOWLEDGMENTS

I frequently hear people saying that doing a PhD is hard. They are right. Asking the right research questions, writing on a review article that never seemed to progress, designing and failing experiments, battling my inner critic, finishing this PhD did not come easy for me. However, when I look back at the past six years, gratitude overwhelms the memories of the struggles I had. During my PhD, I was lucky enough to actively study a topic that I am thoroughly interested in. I met amazing, kind and inspiring people. I developed myself both academically and personally. Importantly, I learned what I love most doing, teaching and supervision, of which I have made my occupation at the time of writing.

Doing a PhD is mostly a solitary journey. Yet, in solitary journeys help and support from other people is maybe more important than anything. In the following I would like to show my gratitude to those ‘other people’ who have helped me on this journey.

First of all, I would like to thank my advisors. Niels, you have a special talent of making people feel energized and inspired after a single meeting. Having a meeting with you after a set-back always made me feel better. Thank you for inspiring me to do novel research and for always being there if I needed your advice.

Marieke, finishing this PhD would not have been possible without your relentless support. Thanks for reading my articles for what seemed the millionth time. For always providing useful feedback and suggestions. Thank you for always staying positive and optimistic in the face of set-backs.

Next to my advisors, I would also like to thank the following people:

My paranymphs and brothers, Xander and Glenn. Thank you for taking my invitation, for being willing to support me at the day of the defense, and for all the time and effort you put into arranging the other important things so that I could finish this booklet and prepare for the defense. You both are incredible!

My reading committee, Jonathan Smallwood, Ritske de Jong and Heleen Slagter. Thank you all for reading my thesis, sharing your comments, and being willing to attend my defense.

Mathanja, co-author on Chapter 3 and collaborator on Chapter 2. I really enjoyed working together with you on getting our experiments to work on the eye tracker, on recruiting participants, and conducting the research itself. Thank you for all the support you offered and for all the feedback and suggestions to improve my work. Thank you

for being willing to share the materials of your Masters thesis with me, which formed the basis of Chapter 3. Without it, Chapter 3 of this thesis would not have been possible.

The CogMod group, Niels, Marieke, Fokie, Jacolien, Jelmer, Katja M., Stephen, Jennifer, Katja P., Oscar, Pallavi, Hermine, Christina, Marlijn, Maarten, Mega, Lionel, Hang, Abby, Corné, and Moritz. Thanks for creating such a supporting and inspiring atmosphere. Thanks for all the useful feedback and suggestions on my research, the inspiring presentations and discussions during the weekly CogMod meetings, and the fun activities we did. I will never forget the - albeit somewhat cringy - acting masterclass we did!

I would also like to thank my fellow AI PhD's. Thank you for the lunches we had together, the fun - sometimes lengthy (not at all my fault ;-)) - discussions on any topic imaginable, for all the fun activities we did (potluck dinners, karting, just to name a few!), for all your feedback on my research, for being faithful pilot participants to my experiments, and so much more! Katja, Oscar, Pallavi, and Moritz, thanks for being such amazing roommates.

Finally, I would like to thank some significant other people:

Pap en mam, bedankt dat jullie altijd in mij geloofd hebben. Voor alle weekenden dat ik bij jullie welkom was. Voor al mijn geklaag dat jullie geduldig verdragen hebben. Zonder jullie onvoorwaardelijke steun in mijn leven was ik nooit tot een PhD gekomen.

Piet en Piety, bedankt dat jullie de deuren van jullie huis voor mij geopend hebben in de eerste maand van mijn PhD project. Ik heb me bij jullie heel erg welkom gevoeld. Piet, bedankt voor het mooie kunstwerk dat nu op de voorkant van het boekje staat te pronken!

Als laatste jou, Dianne. In de laatste fase ben jij ontzettend belangrijk voor mij geweest. Door jouw hulp, steun en rotsvast vertrouwen heb ik mijn proefschrift af kunnen maken. Lief, ik ben je ontzettend dankbaar!

REFERENCES

- Addis, D. R., & Schacter, D. L. (2008). Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus*, 18(2), 227–237. <https://doi.org/10.1002/hipo.20405>
- Allen, M., Smallwood, J., Christensen, J., Gramm, D., Rasmussen, B., Jensen, C. G., ... Lutz, A. (2013). The balanced mind: the variability of task-unrelated thoughts predicts error monitoring. *Frontiers in Human Neuroscience*, 7(November), 1–15. <https://doi.org/10.3389/fnhum.2013.00743>
- Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science* (Vol. 26). [https://doi.org/10.1016/S0364-0213\(01\)00058-1](https://doi.org/10.1016/S0364-0213(01)00058-1)
- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the Environment in Memory. *Psychological Science*, 2(6), 396–408. <https://doi.org/10.1111/j.1467-9280.1991.tb00174.x>
- Andrews-Hanna, J., Smallwood, J., & Spreng, R. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, 1316(1), 29–52. <https://doi.org/10.1111/nyas.12360>
- Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., & Krystal, J. H. (2013). The Role of Default Network Deactivation in Cognition and Disease. *Trends in Cognitive Sciences*, 16(12), 584–592. <https://doi.org/10.1016/j.tics.2012.10.008>
- Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends in Cognitive Sciences*, 21(8), 573–576. <https://doi.org/10.1016/j.tics.2017.05.001>
- Aston-Jones, G., & Cohen, J. D. (2005). An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, 28, 403–450.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition*, 21(2), 146–155. <https://doi.org/10.3758/BF03202726>
- Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, 23(10), 1117–1122. <https://doi.org/10.1177/0956797612446024>
- Baird, B., Smallwood, J., & Schooler, J. W. (2011). Back to the future: Autobiographical planning and the functionality of mind-wandering. *Consciousness and Cognition*, 20(4), 1604–1611. <https://doi.org/10.1016/j.concog.2011.08.007>
- Barlow, D. H. (2002). *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and*

- Panic* (2nd ed.). New York, NY: Guilford Press.
- Barnhofer, T., Crane, C., Spinhoven, P., & Williams, J. M. G. (2007). Failures to retrieve specific memories in previously depressed individuals: Random errors or content-related? *Behaviour Research and Therapy*, 45(8), 1859–1869. <https://doi.org/10.1016/j.brat.2007.02.006>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bastian, M., & Sackur, J. (2013). Mind wandering at the fingertips: Automatic parsing of subjective states based on response time variability. *Frontiers in Psychology*, 4(SEP), 1–11. <https://doi.org/10.3389/fpsyg.2013.00573>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv Preprint: 1506.04967*.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2018). lme4: Linear mixed-effects models using Eigen and S4.
- Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, 194(August 2019), 104057. <https://doi.org/10.1016/j.cognition.2019.104057>
- Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, 91(2), 276–292.
- Beckwé, M., & Deroost, N. (2016). Attentional biases in ruminators and worriers. *Psychological Research*, 80(6), 952–962. <https://doi.org/10.1007/s00426-015-0703-8>
- Benedek, M., Stoiser, R., Walcher, S., & Körner, C. (2017). Eye behavior associated with internally versus externally directed cognition. *Frontiers in Psychology*, 8(JUN), 1–9. <https://doi.org/10.3389/fpsyg.2017.01092>
- Bennike, I. H., Wieghorst, A., & Kirk, U. (2017). Online-based Mindfulness Training Reduces Behavioral Markers of Mind Wandering. *Journal of Cognitive Enhancement*, 1(2), 172–181. <https://doi.org/10.1007/s41465-017-0020-9>
- Berntsen, D. (2019). Spontaneous future cognitions: an integrative review. *Psychological Research*, 83(4), 651–665. <https://doi.org/10.1007/s00426-018-1127-z>
- 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.
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113. <https://doi.org/10.1016/j.tics.2011.12.010>
- Braverman, A., & Meiran, N. (2010). Task conflict effect in task switching. *Psychological Research*, 74(6), 568–578. <https://doi.org/10.1007/s00426-010-0279-2>
- Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil

- diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods*, 45(4), 1322–1331. <https://doi.org/10.3758/s13428-013-0327-0>
- Broadbent, D. E. (1958). *Perception and communication*. London, UK: Pergamon Press.
- Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, 67(6), 380–400. <https://doi.org/10.1037/h0040373>
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, 67, 319–333. <https://doi.org/10.1037/0022-3514.67.2.319>
- Cheyne, A. J., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, 111(1), 98–113. <https://doi.org/10.1016/j.cognition.2008.12.009>
- Christoff, K. (2012). Undirected thought: Neural determinants and correlates. *Brain Research*, 1428, 51–59. <https://doi.org/10.1016/j.brainres.2011.09.060>
- Christoff, K., Gordon, A., & Smith, R. (2011). The role of spontaneous thought in human cognition. In O. Vartanian & D. R. Mandel (Eds.), *Neuroscience of Decision Making* (pp. 259–284). New York, NY, US: Psychology Press.
- Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, N. R., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature Reviews Neuroscience*, 1–44. <https://doi.org/10.1038/nrn.2016.113>
- Cirillo, F. (2018). *The pomodoro technique: The lifechanging time-management system*. Random House.
- Cole, M. W., Bagic, A., Kass, R., & Schneider, W. (2010). Prefrontal Dynamics Underlying Rapid Instructed Task Learning Reverse with Practice. *Journal of Neuroscience*, 30(42), 14245–14254. <https://doi.org/10.1523/JNEUROSCI.1662-10.2010>
- Cole, M. W., Braver, T. S., & Meiran, N. (2017). The task novelty paradox: Flexible control of inflexible neural pathways during rapid instructed task learning. *Neuroscience and Biobehavioral Reviews*, 81, 4–15. <https://doi.org/10.1016/j.neubiorev.2017.02.009>
- Cole, M. W., Ito, T., & Braver, T. S. (2016). The Behavioral Relevance of Task Information in Human Prefrontal Cortex. *Cerebral Cortex*, 26(6), 2497–2505. <https://doi.org/10.1093/cercor/bhvo72>
- Cole, M. W., Laurent, P., & Stocco, A. (2013). Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control. *Cognitive, Affective and Behavioral Neuroscience*, 13(1), 1–22. <https://doi.org/10.3758/s13415-012-0125-7>
- Cole, M. W., Patrick, L. M., Meiran, N., & Braver, T. S. (2018). A role for proactive control

- in rapid instructed task learning. *Acta Psychologica*, 184(March), 20–30. <https://doi.org/10.1016/j.actpsy.2017.06.004>
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786.
- Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, 107(2), 261–288.
- Conway, M. A., Singer, J. A., & Tagini, A. (2004). The Self and Autobiographical Memory: Correspondence and Coherence. *Social Cognition*, 22(5), 491–529. <https://doi.org/10.1521/soco.22.5.491.50768>
- Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, 17(4), 297–338. <https://doi.org/10.1080/026432900380427>
- D'Argembeau, A., Ruby, P., Collette, F., Degueldre, C., Baetee, E., Luxen, A., ... Salmon, E. (2007). Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of Cognitive Neuroscience*, 19(6), 935–944.
- Daamen, J., Vugt, M. K. Van, & Taatgen, N. A. (2016). Measuring and modeling distraction by self-referential processing in a complex working memory task. In *38th Annual Meeting of the Cognitive Science Society*. Philadelphia, PA.
- Dang, J. S., Figueroa, I. J., & Helton, W. S. (2018). You are measuring the decision to be fast, not inattention: the Sustained Attention to Response Task does not measure sustained attention. *Experimental Brain Research*, 236(8), 2255–2262. <https://doi.org/10.1007/s00221-018-5291-6>
- Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655). <https://doi.org/10.1098/rstb.2013.0478>
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711. <https://doi.org/10.1038/nn1560>
- de Dreu, C. K. W., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin*, 38(5), 656–669. <https://doi.org/10.1177/0146167211435795>
- Denson, T. F., Spanovic, M., & Miller, N. (2009). Cognitive Appraisals and Emotions Predict Cortisol and Immune Responses: A Meta-Analysis of Acute Laboratory Social Stressors and Emotion Inductions. *Psychological Bulletin*, 135(6), 823–853. <https://doi.org/10.1037/a0016909>
- Dixon, M. L., Fox, K. C. R., & Christoff, K. (2014). A framework for understanding the

- relationship between externally and internally directed cognition. *Neuropsychologia*, 62(1), 321–330. <https://doi.org/10.1016/j.neuropsychologia.2014.05.024>
- Dohmatob, E., Dumas, G., & Bzdok, D. (2020). Dark control: The default mode network as a reinforcement learning agent. *Human Brain Mapping*, 41(12), 3318–3341. <https://doi.org/10.1002/hbm.25019>
- Duque, A., Sanchez, A., & Vazquez, C. (2014). Gaze-fixation and pupil dilation in the processing of emotional faces: The role of rumination. *Cognition and Emotion*, 28(8), 1347–1366. <https://doi.org/10.1080/02699931.2014.881327>
- Einstein, G. O., & McDaniel, M. A. (1996). Retrieval processes in prospective memory: Theoretical approaches and some new empirical findings. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective Memory: Theory and Applications* (pp. 115–141). Mahwah, NJ: Erlbaum.
- Einstein, G. O., Smith, R. E., McDaniel, M. A., & Shaw, P. (1997). Aging and prospective memory: The influence of increased task demands at encoding and retrieval. *Psychology and Aging*, 12(3), 479–488. <https://doi.org/10.1037/0882-7974.12.3.479>
- Ellamil, M., Fox, K. C. R., Dixon, M. L., Pritchard, S., Todd, R. M., Thompson, E., & Christoff, K. (2016). Dynamics of neural recruitment surrounding the spontaneous arising of thoughts in experienced mindfulness practitioners. *NeuroImage*, 136, 186–196. <https://doi.org/10.1016/j.neuroimage.2016.04.034>
- Faber, M., & Mills, C. (2018). The Critical Role of the Hippocampus in Mind Wandering. *The Journal of Neuroscience*, 38(29), 6439–6441. <https://doi.org/10.1523/JNEUROSCI.0995-18.2018>
- Farmer, G. D., Janssen, C. P., Nguyen, A. T., & Brumby, D. P. (2018). Dividing Attention Between Tasks: Testing Whether Explicit Payoff Functions Elicit Optimal Dual-Task Performance. *Cognitive Science*, 42, 820–849. <https://doi.org/10.1111/cogs.12513>
- Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, 20(3), 586–592.
- Finkbeiner, K. M., Wilson, K. M., Russell, P. N., & Helton, W. S. (2015). The effects of warning cues and attention-capturing stimuli on the sustained attention to response task. *Experimental Brain Research*, 233(4), 1061–1068. <https://doi.org/10.1007/s00221-014-4179-3>
- Forster, S., & Lavie, N. (2009). Harnessing the wandering mind: The role of perceptual load. *Cognition*, 111(3), 345–355. <https://doi.org/10.1016/j.cognition.2009.02.006>
- Foulsham, T., Farley, J., & Kingstone, A. (2013). Mind wandering in sentence reading: decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, 67(1), 51–59. <https://doi.org/10.1037/a0030217>
- Fox, K. C., & Beaty, R. E. (2019). Mind-wandering as creative thinking: neural, psychological, and theoretical considerations. *Current Opinion in Behavioral Sciences*, 27, 123–130.

- <https://doi.org/10.1016/j.cobeha.2018.10.009>
- Fox, K. C. R., Nathan Spreng, R., Ellamil, M., Andrews-Hanna, J. R., & Christoff, K. (2015). The wandering brain: Meta-analysis of functional neuroimaging studies of mind-wandering and related spontaneous thought processes. *NeuroImage*, 111, 611–621. <https://doi.org/10.1016/j.neuroimage.2015.02.039>
- Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: pupillometry of spontaneous thought while reading. *Quarterly Journal of Experimental Psychology*, 66(May 2015), 2289–2294. <https://doi.org/10.1080/17470218.2013.858170>
- Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: Using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin and Review*, 18(5), 992–997. <https://doi.org/10.3758/s13423-011-0109-6>
- Gable, S. L., Hopper, E. A., & Schooler, J. W. (2019). When the Muses Strike: Creative Ideas of Physicists and Writers Routinely Occur During Mind Wandering. *Psychological Science*, 30(3), 396–404. <https://doi.org/10.1177/0956797618820626>
- Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods*, 43(4), 1171–1181. <https://doi.org/10.3758/s13428-011-0109-5>
- Gerjets, P., Scheiter, K., & Schorr, T. (2003). Modeling Processes of Volitional Action Control in Multiple-task Performance: How to Explain Effects of Goal Competition and Task Difficulty on Processing Strategies and Performance with ACT-R. *Cognitive Science Quarterly*, 3, 335–400.
- Gershman, S. J., Markman, A. B., & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: a tale of two systems. *Journal of Experimental Psychology: General*, 143(1), 182–194. <https://doi.org/10.1037/a0030844>
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory. *Cognitive, Affective, & Behavioral Neuroscience*, 10(2), 252–269.
- Godden, D., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments. *British Journal of Psychology*, 66, 325–331.
- Godijn, R., & Theeuwes, J. (2012). Overt is no better than covert when rehearsing visuo-spatial information in working memory. *Memory & Cognition*, 40(1), 52–61. <https://doi.org/10.3758/s13421-011-0132-x>
- Gollwitzer, P. M. (1996). The Volitional Benefits of Planning. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 287–312). London, UK: The Guilford Press.
- Goschke, T., & Kuhl, J. (1993). Representation of intentions: Persisting activation in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(5), 1211–1226. <https://doi.org/10.1037/0278-7393.19.5.1211>

- Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in Psychology*, 5(FEB), 1–10.
- Hamilton, J. P., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., & Gotlib, I. H. (2012). Functional Neuroimaging of Major Depressive Disorder: A Meta-Analysis and New Integration of Baseline Activation and Neural Response Data. *The American Journal of Psychiatry*, 169(7), 693–703.
- Harrison, N. A., Singer, T., Rotshtein, P., Dolan, R. J., & Critchley, H. D. (2006). Pupillary contagion: central mechanisms engaged in sadness processing. *Social Cognitive and Affective Neuroscience*, 1(1), 5–17. <https://doi.org/10.1093/scan/nsloo6>
- Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, 11(7), 299–306. <https://doi.org/10.1016/j.tics.2007.05.001>
- Hiatt, L. M., & Trafton, G. J. (2015). A Computational Model of Mind Wandering. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *In Proceedings 37th Annual Meeting of the Cognitive Science Society* (pp. 914–919). Austin, TX: Cognitive Science Society.
- Hoeks, B., & Levelt, W. J. M. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments, & Computers*, 25(1), 16–26. <https://doi.org/10.3758/BF03204445>
- Huijser, S., Taatgen, N. A., Vugt, M. K. Van, & van Vugt, M. K. (2021). The art of planning ahead: When do we prepare for the future and when is it effective? *Journal of Experimental Psychology: Learning Memory and Cognition*, 47(5), 705–726. <https://doi.org/10.1037/xlm0000970>
- Huijser, S., van Vugt, M. K. M. K., & Taatgen, N. A. N. A. (2018). The wandering self: Tracking distracting self-generated thought in a cognitively demanding context. *Consciousness and Cognition*, 58, 170–185. <https://doi.org/10.1016/j.concog.2017.12.004>
- Janssen, C. P., & Brumby, D. P. (2015). Strategic adaptation to task characteristics, incentives, and individual differences in dual-tasking. *PLoS ONE*, 10(7), 1–33. <https://doi.org/10.1371/journal.pone.0130009>
- Janssen, C. P., Everaert, E., Hendriksen, H. M. A., Mensing, G. L., Tigchelaar, L. J., & Nunner, H. (2019). The influence of rewards on (sub-)optimal interleaving. *PLoS ONE*, 14(3), 1–25. <https://doi.org/10.1371/journal.pone.0214027>
- Jin, C. Y., Borst, J. P., & van Vugt, M. K. (2019). Predicting task-general mind-wandering with EEG. *Cognitive, Affective and Behavioral Neuroscience*, 19, 1059–1073.
- Jin, C. Y., Borst, J. P., & van Vugt, M. K. (2020). Distinguishing vigilance decrement and low task demands from mind-wandering: A machine learning analysis of EEG. *European Journal of Neuroscience*, 52(9), 4147–4164. <https://doi.org/10.1111/ejn.14863>

- Joormann, J., Levens, S. M., & Gotlib, I. H. (2011). Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological Science*, 22(8), 979–983. <https://doi.org/10.1177/0956797611415539>
- Jubera-García, E., Gevers, W., & Opstal, F. Van. (2019). Influence of content and intensity of thought on behavioral and pupil changes during active mind-wandering , off-focus , and on-task states. *Attention, Perception, & Psychophysics*, 1–11.
- Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702), 1776–1780.
- Kam, J. W. Y., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow Fluctuations in Attentional Control of Sensory Cortex. *Journal of Cognitive Neuroscience*, 23(2), 460–470. <https://doi.org/http://dx.doi.org/10.1162/jocn.2010.21443>
- Kam, J. W. Y., & Handy, T. C. (2013). The neurocognitive consequences of the wandering mind: a mechanistic account of sensory-motor decoupling. *Frontiers in Psychology*, 4(October), 725. <https://doi.org/10.3389/fpsyg.2013.00725>
- Kane, M. J., Gross, G. M., Chun, C. A., Smeekens, B. A., Meier, M. E., Silvia, P. J., & Kwapil, T. R. (2017). For Whom the Mind Wanders, and When, Varies Across Laboratory and Daily-Life Settings. *Psychological Science*, 28(9), 1271–1289. <https://doi.org/10.1177/0956797617706086>
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131(1), 66–71.
- Katidioti, I., & Taatgen, N. A. (2014). Choice in Multitasking. *Human Factors*, 56(4), 728–736.
- Kavanagh, D. J., Andrade, J., & May, J. (2005). Imaginary Relish and Exquisite Torture: The Elaborated Intrusion Theory of Desire David. *Psychological Review*, 112(2), 446–467.
- Keele, S. W. (1973). *Attention and human performance*. Pacific Palisades, CA: Goodyear Publishing Company.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650. <https://doi.org/10.3758/BRM.42.3.643>
- Kieras, D. E. (2017a). A Summary of the EPIC Cognitive Architecture. In S. E. F. Chipman (Ed.), *The Oxford Handbook of Cognitive Science* (pp. 27–48). New York, NY: Oxford University Press.
- Kieras, D. E. (2017b). EPIC lessons for the proposed standard model of the mind. In *AAAI Fall Symposium - Technical Report* (pp. 388–393).

- Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, 330(6006), 932–932.
- Kliegel, M., Mackinlay, R., & Jäger, T. (2008). Complex Prospective Memory : Development Across the Lifespan and the Role of Task Interruption. *Developmental Psychology*, 44(2), 612–617. <https://doi.org/10.1037/0012-1649.44.2.612>
- Kliegel, M., McDaniel, M. A., & Einstein, G. O. (2000). Plan formation , retention , and execution in prospective memory : A new approach and age-related effects. *Memory & Cognition*, 28(6), 1041–1049.
- Kliegel, M., McDaniel, M. A., Einstein, G. O., & Moor, C. (2007). Realizing complex delayed intentions in young and old adults : The role of planning aids. *Memory & Cognition*, 35(7), 1735–1746.
- Klinger, E. (2013). Goal commitments and the content of thoughts and dreams: Basic principles. *Frontiers in Psychology*, 4(JUL), 1–17.
- Klinger, E., Barta, S. G., & Maxeiner, M. E. (1980). Motivational correlates of thought content frequency and commitment. *Journal of Personality and Social Psychology*, 39(6), 1222–1237.
- Klinger, E., & Cox, W. M. (2004). Motivation and the Goal Theory of Current Concerns. In W. M. Cox & E. Klinger (Eds.), *Handbook of Motivational Counseling: Goal-Based Approaches to Assessment and Intervention with Addiction and Other Problems* (pp. 3–29). Chichester: John Wiley & Sons.
- Klinger, E., & Miles Cox, W. (2011). Motivation and the Theory of Current Concerns. *Handbook of Motivational Counseling: Concepts, Approaches, and Assessment*. <https://doi.org/10.1002/9780470713129.ch1>
- Konishi, M., Brown, K., Battaglini, L., & Smallwood, J. (2017). When attention wanders: Pupillometric signatures of fluctuations in external attention. *Cognition*, 168, 16–26. <https://doi.org/10.1016/j.cognition.2017.06.006>
- Konishi, M., & Smallwood, J. (2016). Shadowing the wandering mind: how understanding the mind-wandering state can inform our appreciation of conscious experience. *Wiley Interdisciplinary Reviews: Cognitive Science*. <https://doi.org/10.1002/wcs.1392>
- Koster, E. H. W., De Lissnyder, E., & De Raedt, R. (2013). Rumination is characterized by valence-specific impairments in switching of attention. *Acta Psychologica*, 144(3), 563–570. <https://doi.org/10.1016/j.actpsy.2013.09.008>
- Kvavilashvili, L., & Fisher, L. (2007). Is time-based prospective remembering mediated by self-initiated rehearsals? Role of incidental cues, ongoing activity, age, and motivation. *Journal of Experimental Psychology: General*, 136(1), 112–132. <https://doi.org/10.1037/0096-3445.136.1.112>
- Lebiere, C., & Lee, F. J. (2002). Intention superiority effect: A context-switching account. *Cognitive Systems Research*, 3(1), 57–65. [https://doi.org/10.1016/S1389-0417\(01\)00044-4](https://doi.org/10.1016/S1389-0417(01)00044-4)

- Leszczynski, M., Chaieb, L., Reber, T. P., Derner, M., Axmacher, N., & Fell, J. (2017). Mind wandering simultaneously prolongs reactions and promotes creative incubation. *Scientific Reports*, 7(1), 10197. <https://doi.org/10.1038/s41598-017-10616-3>
- Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The Persistence of Thought: Evidence for a Role of Working Memory in the Maintenance of Task-Unrelated Thinking. *Psychological Science*, 23(4), 375–380. <https://doi.org/10.1177/0956797611431465>
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hillsdale, NJ, NJ: Erlbaum.
- Macken, W. J. (2002). Environmental Context and Recognition : The Role of Recollection and Familiarity. *Journal Experimental Psychology: Learning, Memory and Cognition*, 28(1), 153–161. <https://doi.org/10.1037//0278-7393.28.1.153>
- Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia*, 37(6), 661–670. [https://doi.org/10.1016/S0028-3932\(98\)00127-4](https://doi.org/10.1016/S0028-3932(98)00127-4)
- Marchetti, I., Koster, E. H. W., Klinger, E., & Alloy, L. B. (2016). Spontaneous thought and vulnerability to mood disorders: The dark side of the wandering mind. *Clinical Psychological Science*, 4(5), 835–857. <https://doi.org/10.1177/2167702615622383>
- Marsh, R. L., Hancock, T. W., & Hicks, J. L. (2002). The demands of an ongoing activity influence the success of event-based prospective memory. *Psychonomic Bulletin and Review*, 9(3), 604–610. <https://doi.org/10.3758/BF03196319>
- Marsh, R. L., Hicks, J. L., & Bryan, E. S. (1999). The activation of unrelated and canceled intentions. *Memory & Cognition*, 27(2), 320–327. <https://doi.org/10.3758/BF03211415>
- Martin, J. G. A., Nussey, D. H., Wilson, A. J., & Réale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods in Ecology and Evolution*, 2(4), 362–374. <https://doi.org/10.1111/j.2041-210X.2010.00084.x>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed effects models. *Journal of Memory and Language*, 64, 305–315.
- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1124–1140. <https://doi.org/10.1037//0278-7393.26.5.1124>
- McCrea, S. M., Penningroth, S. L., Radakovich, M. P., Mccrea, S. M., Penningroth, S. L., & Radakovich, M. P. (2017). Implementation intentions forge a strong cue – response link and boost prospective memory performance Implementation intentions forge a strong cue – response link and boost prospective memory performance. *Journal of Cognitive Psychology*, 27(1), 12–26. <https://doi.org/10.1080/20445911.2014.975816>
- McDaniel, M. A., Howard, D. C., & Butler, K. M. (2008). Implementation intentions facilitate prospective memory under high attention demands. *Memory & Cognition*, 36(4),

- 716–724. <https://doi.org/10.3758/MC.36.4.716>
- McDaniel, M. A., & Scullin, M. K. (2010). Implementation intention encoding does not automatize prospective memory responding. *Memory*, 38(2), 221–232. <https://doi.org/10.3758/MC.38.2.221>
- McDaniel, M. A., Umanath, S., Einstein, G. O., & Waldum, E. R. (2015). Dual pathways to prospective remembering. *Frontiers in Human Neuroscience*, 9(July), 1–12. <https://doi.org/10.3389/fnhum.2015.00392>
- McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension. *Journal of Experimental Psychology: General*, 141(2), 302–320.
- McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1), 196–204.
- McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin*, 136(2), 188–197. <https://doi.org/10.1037/a0018298>
- McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “D’oh!”: working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 525–549.
- McVay, J. C., & Kane, M. J. (2013). Dispatching the wandering mind? Toward a laboratory method for cuing “spontaneous” off-task thought. *Frontiers in Psychology*, 4(SEP), 1–16.
- Medea, B., Karapanagiotidis, T., Konishi, M., Ottaviani, C., Margulies, D., Bernasconi, A., ... Smallwood, J. (2016). How do we decide what to do? Resting-state connectivity patterns and components of self-generated thought linked to the development of more concrete personal goals. *Experimental Brain Research*, 1–13. <https://doi.org/10.1007/s00221-016-4729-y>
- Meilán, J. J. G., Carro, J., Arana, J. M., & Pérez, E. (2011). Intention superiority effect on implicit memory of shopping lists: Activation and inhibition processes. *Motivation and Emotion*, 35(2), 144–150. <https://doi.org/10.1007/s11031-011-9205-1>
- Meyer, D. E., & Kieras, D. E. (1997). A Computational Theory of Executive Cognitive Processes and Multiple-Task Performance: Part 1: Basic MEchanisms. *Psychological Review*, 104(1), 3–65. <https://doi.org/10.1037/0033-295X.104.1.3>
- Mills, C., Herrera-Bennett, A., Faber, M., & Christoff, K. (2018). Why the mind wanders: How spontaneous thought’s default variability may support episodic efficiency and semantic optimization. In K. C. R. Fox & K. Christoff (Eds.), *The Oxford Handbook of Spontaneous Thought: Mind-Wandering, Creativity, and Dreaming* (pp. 11–22). New York: Oxford University Press.

- Mills, C., Raffaelli, Q., Irving, Z. C., Stan, D., & Christoff, K. (2018). Is an off-task mind a freely-moving mind? Examining the relationship between different dimensions of thought. *Consciousness and Cognition*, 58(May 2017), 20–33. <https://doi.org/10.1016/j.concog.2017.10.003>
- Mittner, M., Boekel, W., Tucker, a. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the Brain Takes a Break: A Model-Based Analysis of Mind Wandering. *Journal of Neuroscience*, 34(July), 16286–16295.
- Mittner, Matthias, Hawkins, G. E., Boekel, W., & Forstmann, B. U. (2016). A Neural Model of Mind Wandering. *Trends in Cognitive Sciences*, 20(8), 570–578. <https://doi.org/10.1016/j.tics.2016.06.004>
- Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *ELife*, 7, 1–25. <https://doi.org/10.7554/eLife.32548>
- Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 67(1), 11–18. <https://doi.org/10.1037/a0031569>
- Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Current Eye Research*, 21(1), 535–542. [https://doi.org/10.1076/0271-3683\(200007\)2111-ZFT535](https://doi.org/10.1076/0271-3683(200007)2111-ZFT535)
- Morris, R., & Ward, G. (2005). *The Cognitive Psychology of Planning*. Hove, UK: Psychology Press.
- Moye, A. J., & Van Vugt, M. K. (2021). A Computational Model of Focused Attention Meditation and Its Transfer to a Sustained Attention Task. *IEEE Transactions on Affective Computing*, 12(2), 329–339. <https://doi.org/10.1109/TAFFC.2019.2908172>
- Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness Training Improves Working Memory Capacity and GRE Performance While Reducing Mind Wandering. *Psychological Science*, 24(5), 776–781. <https://doi.org/10.1177/0956797612459659>
- Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General*, 141(4), 788–798.
- Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2012). Mindfulness and mind-wandering: Finding convergence through opposing constructs. *Emotion*, 12(3), 442–448. <https://doi.org/10.1037/a0026678>
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5), 482–488. <https://doi.org/10.1037/h0045106>
- Murphy, P. R., O'Connell, R. G., O' Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, 35(8), 4140–4154.

- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and p3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, 48(11), 1532–1543.
- Nijboer, M., Borst, J. P., van Rijn, H., & Taatgen, N. A. (2016). Contrasting Single and Multi-Component Working-Memory Systems in Dual Tasking. *Cognitive Psychology*, 1–26.
- 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). <https://doi.org/10.1371/journal.pone.0079583>
- Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, 109(3), 504–511. <https://doi.org/10.1037/0021-843X.109.3.504>
- Nolen-Hoeksema, Susan, & Morrow, J. (1991). A prospective study of depression and distress following a natural disaster: The 1989 Loma Prieta earthquake. *Journal of Personality and Social Psychology*, 61(1), 105–121.
- Nolen-Hoeksema, Susan, Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking Rumination. *Perspectives on Psychological Science*, 3(5), 400–424. <https://doi.org/10.1111/j.1745-6924.2008.00088.x>
- Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, 134(3), 368–387. <https://doi.org/10.1037/0096-3445.134.3.368>
- Oberauer, K., & Hein, L. (2012). Attention to Information in Working Memory. *Current Directions in Psychological Science*, 21(3), 164–169. <https://doi.org/10.1177/0963721412444727>
- Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology*, 66(2), 157–211. <https://doi.org/10.1016/j.cogpsych.2012.11.001>
- Ottaviani, C., Shapiro, D., & Couyoumdjian, A. (2013). Flexibility as the key for somatic health: From mind wandering to perseverative cognition. *Biological Psychology*, 94(1), 38–43. <https://doi.org/10.1016/j.biopsycho.2013.05.003>
- Ottaviani, C., Shapiro, D., & Fitzgerald, L. (2011). Rumination in the laboratory: What happens when you go back to everyday life? *Psychophysiology*, 48(4), 453–461. <https://doi.org/10.1111/j.1469-8986.2010.01122.x>
- Papies, E., Stroebe, W., & Aarts, H. (2007). Pleasure in the mind: Restrained eating and spontaneous hedonic thoughts about food. *Journal of Experimental Social Psychology*, 43(5), 810–817. <https://doi.org/10.1016/j.jesp.2006.08.001>
- Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 358–377. <https://doi.org/10.1037/0096-1523.10.3.358>

- Pashler, H. (1994). Dual-Task Interference in Simple Tasks: Data and Theory. *Psychological Bulletin*, 116(2), 220–244. <https://doi.org/10.1037/0033-2909.116.2.220>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Peirce, J. W. (2007). Psychopy - psychophysics software in python. *Journal of Neuroscience Methods*, 162(1), 8–13.
- Pereg, M., & Meiran, N. (2019). Rapid instructed task learning (but not automatic effects of instructions) is influenced by working memory load. *PLoS ONE*, 14(6), 1–25. <https://doi.org/10.1371/journal.pone.0217681>
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, 28(3), 164–171.
- Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye Movements During Mindless Reading. *Psychological Science*, 21(9), 1300–1310. <https://doi.org/10.1177/0956797610378686>
- Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). “Oops!”: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35(6), 747–758.
- Robison, M. K., Miller, A. L., & Unsworth, N. (2019). Examining the effects of probe frequency, response options, and framing within the thought-probe method. *Behavior Research Methods*, 51(1), 398–408. <https://doi.org/10.3758/s13428-019-01212-6>
- Rummel, J., & Boywitt, C. D. (2014). Controlling the stream of thought: Working memory capacity predicts adjustment of mind-wandering to situational demands. *Psychonomic Bulletin & Review*, 21(5), 1309–1315.
- Rummel, J., Einstein, G. O., & Rampey, H. (2012). Implementation-intention encoding in a prospective memory task enhances spontaneous retrieval of intentions. *Memory*, 20(8), 803–817. <https://doi.org/10.1080/09658211.2012.707214>
- Salvucci, D., & Bogunovich, P. (2010). Multitasking and monotasking: The effects of mental workload on deferred task interruptions. In *CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 85–88). Retrieved from <http://dl.acm.org/citation.cfm?id=1753340>
- Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review*, 115(1), 101–130.
- Salvucci, D. D., & Taatgen, N. A. (2010). *The multitasking mind*. Oxford University Press.
- Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption. In *Proceedings of the SIGCHI Conference on*

- Human Factors in Computing Systems: CHI 2009* (pp. 1819–1828). <https://doi.org/10.1145/1518701.1518981>
- Schacter, D. L., & Addis, D. R. (2009). On the nature of medial temporal lobe contributions to the constructive simulation of future events Email alerting service On the nature of medial temporal lobe contributions to the constructive simulation of future events. *Society*, 1245–1253. <https://doi.org/10.1098/rstb.2008.0308>
- Schacter, D. L., Addis, D. R., & Szpunar, K. K. (2017). Escaping the Past: Contributions of the Hippocampus to Future Thinking and Imagination. In D. E. Hannula & M. C. Duff (Eds.), *The Hippocampus from Cells to Systems* (pp. 439–465). Springer International Publishing. <https://doi.org/10.1007/978-3-319-50406-3>
- Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, 49(1), 289–318.
- Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences*, 6(8), 339–344.
- Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, 15(7), 319–326. <https://doi.org/10.1016/j.tics.2011.05.006>
- Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology*, 32(3), 219–250. <https://doi.org/10.1006/cogp.1997.0652>
- Schuck, N. W., & Niv, Y. (2019). Sequential replay of nonspatial task states in the human hippocampus. *Science*, 364(6447). <https://doi.org/10.1126/science.aaw5181>
- Schult, J. C., & Steffens, M. C. (2017). The effects of enactment and intention accessibility on prospective memory performance. *Memory & Cognition*, 45(4), 625–638. <https://doi.org/10.3758/s13421-016-0677-9>
- Schwarz, N. (2007). Retrospective and Concurrent Self-Reports: The Rationale for Real-Time Data Capture. In *The science of real-time data capture: Self-reports in health research* (pp. 11–26). New York, NY, US: Oxford University Press.
- Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies*, 4(1), 5–34. <https://doi.org/10.1023/a:1023605205115>
- Scullin, M. K., Bugg, J. M., McDaniel, M. A., & Einstein, G. O. (2011). Prospective memory and aging: Preserved spontaneous retrieval, but impaired deactivation, in older adults. *Memory & Cognition*, 39(7), 1232–1240. <https://doi.org/10.3758/s13421-011-0106-z>
- Scullin, M. K., McDaniel, M. A., Dasse, M. N., Lee, J., Kurinec, A., Tami, C., & Krueger, M. L. (2018). Thought probes during prospective memory encoding : Evidence for perfunctory processes. *PLoS ONE*, 13(6), 1–26.
- Seli, P., Carriere, J. S. A., Wammes, J. D., Risko, E. F., Schacter, D. L., & Smilek, D. (2018). On

- the Clock: Evidence for the Rapid and Strategic Modulation of Mind Wandering. *Psychological Science*, 1–10.
- Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 1–5. <https://doi.org/10.1037/a0030954>
- Seli, P., Cheyne, J. A., Xu, M., Purdon, C., Smilek, D., Seli, P., ... Smilek, D. (2015). Motivation, Intentionality, and Mind Wandering: Implications for Assessments of Task-Unrelated Thought. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(5), 1417–1425.
- Seli, P., Konishi, M., Risko, E. F., & Smilek, D. (2018). The role of task difficulty in theoretical accounts of mind wandering. *Consciousness and Cognition*, 65(September), 255–262. <https://doi.org/10.1016/j.concog.2018.08.005>
- Seli, P., Ralph, B. C. W., Risko, E. F., W. Schooler, J., Schacter, D. L., & Smilek, D. (2017). Intentionality and meta-awareness of mind wandering: Are they one and the same, or distinct dimensions? *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-017-1249-0>
- Seli, P., Risko, E. F., & Smilek, D. (2016). On the Necessity of Distinguishing Between Unintentional and Intentional Mind Wandering. *Psychological Science*, 27(5), 685–691.
- Shohamy, D., & Daw, N. D. (2015). Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences*, 5, 85–90. <https://doi.org/10.1016/j.cobeha.2015.08.010>
- Siegle, G. J., Steinhauer, S. R., Carter, C. S., Ramel, W., & Thase, M. E. (2003). Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self-reported rumination. *Cognitive Therapy and Research*, 27(3), 365–382. <https://doi.org/10.1023/A:1023974602357>
- Smallwood, J. (2013). Distinguishing how from why the mind wanders: a process-occurrence framework for self-generated mental activity. *Psychological Bulletin*, 139(3), 519–535. <https://doi.org/10.1037/a0030010>
- Smallwood, J., & Andrews-Hanna, J. (2013). Not all minds that wander are lost: The importance of a balanced perspective on the mind-wandering state. *Frontiers in Psychology*, 4(AUG), 1–6.
- Smallwood, J., Brown, K. S., Baird, B., Mrazek, M. D., Franklin, M. S., & Schooler, J. W. (2012). Insulation for daydreams: A role for tonic norepinephrine in the facilitation of internally guided thought. *PLoS ONE*, 7(4), 1–5. <https://doi.org/10.1371/journal.pone.0033706>
- Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., ... Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PLoS ONE*, 6(3). <https://doi.org/10.1371/journal.pone.0018298>

- Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, 13(4), 657–690. <https://doi.org/10.1016/j.concog.2004.06.003>
- Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: the curious incident of the wandering mind. *Memory & Cognition*, 36(6), 1144–1150. <https://doi.org/10.3758/MC.36.6.1144>
- Smallwood, J., Nind, L., & Connor, R. C. O. (2009). When is your head at? An exploration of the factors associated with the temporal focus of the wandering mind. *Consciousness and Cognition*, 18(1), 118–125. <https://doi.org/10.1016/j.concog.2008.11.004>
- Smallwood, J., Ruby, F. J. M., & Singer, T. (2013). Letting go of the present: Mind-wandering is associated with reduced delay discounting. *Consciousness and Cognition*, 22(1), 1–7. <https://doi.org/10.1016/j.concog.2012.10.007>
- Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, 132(6), 946–958. <https://doi.org/10.1037/0033-2909.132.6.946>
- Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, 66(1), 487–518.
- Smallwood, J., Schooler, J. W., Turk, D. J., Cunningham, S. J., Burns, P., & Macrae, C. N. (2011). Self-reflection and the temporal focus of the wandering mind. *Consciousness and Cognition*, 20(4), 1120–1126.
- Smith, S., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, 8(2), 203–220.
- Spreng, N. R., Gerlach, K. D., Turner, G. R., & Schacter, D. L. (2015). Autobiographical planning and the brain: Activation and its modulation by qualitative features. *Journal of Cognitive Neuroscience*, 27(11), 2147–2157. <https://doi.org/10.1162/jocn>
- Spreng, N. R., Stevens, D. W., Chamberlain, J. P., Gilmore, A. W., & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage*, 53(1), 303–317. <https://doi.org/10.1016/j.neuroimage.2010.06.016>. Default
- Spronken, M., Holland, R. W., Figner, B., & Dijksterhuis, A. (2016). Temporal focus, temporal distance, and mind-wandering valence: Results from an experience sampling and an experimental study. *Consciousness and Cognition*, 41, 104–118. Retrieved from <http://dx.doi.org/10.1016/j.concog.2016.02.004>
- Stawarczyk, D., Cassol, H., Argembeau, A. D., Smallwood, J., Planck, M., Ruby, F. J. M., & Planck, M. (2013). Phenomenology of future-oriented mind-wandering episodes. *Frontiers in Psychology*, 4(July), 1–12. <https://doi.org/10.3389/fpsyg.2013.00425>
- Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-

- wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, 136(3), 370–381. <https://doi.org/10.1016/j.actpsy.2011.01.002>
- Steindorf, L., & Rummel, J. (2017). “I Should not Forget the Apples!”-Mind-Wandering Episodes Used as Opportunities for Rehearsal in an Interrupted Recall Paradigm. *Applied Cognitive Psychology*, 31(4), 424–430. <https://doi.org/10.1002/acp.3328>
- Szaras, K., & Niedźwieńska, A. (2011). The role of rehearsals in self-generated prospective memory tasks. *International Journal of Psychology*, 46(5), 346–353. <https://doi.org/10.1080/00207594.2011.565342>
- Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, 120(3), 439–471.
- Taatgen, N. A., van Vugt, M. K., Daamen, J., Katidioti, I., Huijser, S., & Borst, J. P. (2021). The resource-availability model of distraction and mind-wandering. *Cognitive Systems Research*, 68, 84–104. <https://doi.org/10.1016/j.cogsys.2021.03.001>
- Tan, T., Zou, H., Chen, C., & Luo, J. (2015). Mind Wandering and the Incubation Effect in Insight Problem Solving. *Creativity Research Journal*, 27(4), 375–382. <https://doi.org/10.1080/10400419.2015.1088290>
- Teasdale, J. D., Proctor, L., Lloyd, C. A., & Baddeley, A. D. (1993). Working memory and stimulus-independent thought: Effects of memory load and presentation rate. *European Journal of Cognitive Psychology*, 5(4), 417–433.
- Thomson, D. R., Besner, D., & Smilek, D. (2013). In pursuit of off-task thought : mind wandering-performance trade-offs while reading aloud and color naming. *Frontiers in Psychology*, 4(June), 1–11.
- Unsworth, N., & Robison, M. K. (2014). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic Bulletin & Review*, 22(3), 757–765.
- Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, 16(4), 601–615.
- Uzzaman, S., & Joordens, S. (2011). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition*, 20(4), 1882–1886. <https://doi.org/10.1016/j.concog.2011.09.010>
- Van den Brink, R. L., Murphy, P. R., & Nieuwenhuis, S. (2016). Pupil diameter tracks lapses of attention. *PLoS ONE*, 11(10), 1–16. <https://doi.org/10.1371/journal.pone.0165274>
- van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, 23, 1–22. <https://doi.org/10.1177/2331216519832483>
- van Rij, J., Wieling, M., Baayen, R. H., & Van Rijn, H. (2017). itsadug: Interpreting Time Series and Autocorrelated Data using GAMMs. Retrieved from <https://cran.r-project.org/web/packages/itsadug>

- van Vugt, M. K., & Broers, N. (2016). Self-reported stickiness of mind-wandering affects task performance. *Frontiers in Psychology*, 7(MAY), 1–8. <https://doi.org/10.3389/fpsyg.2016.00732>
- van Vugt, M. K., Taatgen, N. A., Sackur, J., & Bastian, M. (2015). Mind-wandering due to goal competition : a mechanistic model. In N. Taagen, M. van Vugt, B. J, & K. Melhorn (Eds.), *Proceedings of the 13th International Conference on Cognitive Modeling* (pp. 252–257). University of Groningen.
- Vanucci, M., & Agnoli, S. (2019). Thought Dynamics: Which Role for Mind Wandering in Creativity? In R. Beghetto & G. Corazza (Eds.), *Dynamic Perspectives on Creativity* (pp. 245–260). Cham: Springer.
- Verkuil, B., Brosschot, J., Gebhardt, W., & Thayer, J. (2010). When Worries Make You Sick: A Review of Perseverative Cognition, the Default Stress Response and Somatic Health. *Journal of Experimental Psychopathology*, 1(1), 87–118. <https://doi.org/10.5127/jep.009110>
- Verplanken, B., Friborg, O., Wang, C. E., Trafimow, D., & Woolf, K. (2007). Mental habits: Metacognitive reflection on negative self-thinking. *Journal of Personality and Social Psychology*, 92(3), 526–541. <https://doi.org/10.1037/0022-3514.92.3.526>
- Wallace, B. A. (1999). The Buddhist tradition of Samatha: Methods for refining and examining consciousness. *Journal of Consciousness Studies*, 6(3), 175–187.
- Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behaviour Research Methods*, 50(2), 642–661. <https://doi.org/10.3758/s13428-017-0891-9>
- Wessel, I., Postma, I. R., Huntjens, R. J. C., Crane, C., Smets, J., Zeeman, G. G., & Barnhofer, T. (2014). Differential correlates of autobiographical memory specificity to affective and self-discrepant cues. *Memory*, 22(6), 655–668.
- Whitfield-Gabrieli, S., & Ford, J. M. (2012). Default Mode Network Activity and Connectivity in Psychopathology. *Annual Review of Clinical Psychology*, 8(1), 49–76. <https://doi.org/10.1146/annurev-clinpsy-032511-143049>
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159–177. <https://doi.org/10.1080/14639220210123806>
- Wierda, S. M., Van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22), 8456–8460. <https://doi.org/10.1073/pnas.1201858109>
- Wilhelm, B., Giedke, H., Lüdtke, H., Bittner, E., Hofmann, A., & Wilhelm, H. (2001). Daytime variations in central nervous system activation measured by a pupillographic sleepiness test. *Journal of Sleep Research*, 10(1), 1–7. <https://doi.org/10.1046/j.1365-2869.2001.00239.x>
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood

- estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society*, 73(1), 3–36. <https://doi.org/10.1111/j.14679868.2010.00749.x>
- Wood, S. N. (2017a). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC. <https://doi.org/10.1201/978131537027>
- Wood, S. N. (2017b). mgcv. Mixed gam computation vehicle with automatic smoothness estimation. Retrieved from <https://cran.r-project.org/web/packages/mgcv>
- Wood, S. N., Goude, Y., & Shaw, S. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 64(1), 139–155. <https://doi.org/10.1111/rssc.12068>
- Wood, S. N., Pya, N., & Saefken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of American Statistical Association*, 111, 1548–1575.
- Xu, M., Purdon, C., Seli, P., & Smilek, D. (2017). Mindfulness and mind wandering: The protective effects of brief meditation in anxious individuals. *Consciousness and Cognition*, 51(April), 157–165. <https://doi.org/10.1016/j.concog.2017.03.009>

LIST OF PUBLICATIONS

Journal papers

- Taatgen, N. A., van Vugt, M. K., Daamen, J., Katidioti, I., **Huijser, S.**, & Borst, J. P. (2021). The resource-availability model of distraction and mind-wandering. *Cognitive Systems Research*, 68, 84-104.
- Huijser, S.**, Taatgen, N. A., & van Vugt, M. K. (2021). The art of planning ahead: When do we prepare for the future and when is it effective? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(5), 705-726. <https://doi.org/10.1037/xlm0000970>
- Huijser, S.**, Verkaik, M., van Vugt, M. K., & Taatgen, N. A. (2020). Captivated by thought: “Sticky” thinking leaves traces of perceptual decoupling in task-evoked pupil size. *PLoS ONE*, 15(12), e0243532.
- Huijser, S.**, van Vugt, M. K., & Taatgen, N. A. (2018). The wandering self: Tracking distracting self-generated thought in a cognitively demanding context. *Consciousness and Cognition*, 58, 170-185.

Conference papers

- Huijser, S.**, Taatgen, N., & van Vugt, M. (2017, November). Distracted in a demanding task: A classification study with artificial neural networks. In 29th Benelux Conference on Artificial Intelligence (p. 199-212). Groningen, Netherlands.

Conference abstracts

- Huijser, S.**, Taatgen, N., & van Vugt, M. (2016, August). Following the Wandering Mind in the Eyes: Tracking Distraction by the Self in a Complex Working Memory Task. In D. Reitter, & F. E. Ritter (Eds.), *Proceedings of ICCM 2016 - 14th International Conference on Cognitive Modeling* (pp. 258-260). The Pennsylvania State University.