university of groningen

University Medical Center Groningen

University of Groningen

The problem state bottleneck

Borst, Jelmer

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:
2012

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Borst, J. (2012). *The problem state bottleneck: modeling the behavoral and neural signatures of a cognitive bottleneck in human multitasking*. s.n.

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

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

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

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

Download date: 21-01-2023

The Problem State Bottleneck

Modeling the Behavioral and Neural
Signatures of a Cognitive Bottleneck
in Human Multitasking

Jelmer Pieter Borst

Printed by Gildeprint Drukkerijen,
Enschede, the Netherlands.

MIX
Paper from
responsible sources
FSC® C012854

ISBN printed version: 978-90-367-5286-2
ISBN digital version: 978-90-367-5285-5

© Jelmer P. Borst, Groningen, the Netherlands, 2011.

RIJKSUNIVERSITEIT GRONINGEN

The Problem State Bottleneck

Modeling the Behavioral and Neural Signatures of a Cognitive Bottleneck in Human Multitasking

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
vrijdag 20 april 2012
om 14:30 uur

door

Jelmer Pieter Borst

geboren op 15 december 1983
te Tietjerksteradeel

Promotor:	Prof. dr. N.A. Taatgen
Copromotor:	Dr. H. van Rijn
Beoordelingscommissie:	Prof. dr. R. de Jong
	Prof. dr. J.G.W. Raaijmakers
	Prof. dr. D.D. Salvucci

Contents

Introduction

In which we give a short overview of this dissertation and discuss the underlying theories and applied methodologies.

Parts of this chapter were adapted from Chapters 2, 4, & 6 of this dissertation, and from:

Borst, J.P., Taatgen, N.A., & Van Rijn, H. (in preparation). Using Cognitive Architectures to Analyze fMRI Data of Complex Tasks. In A. Johnson & R. W. Proctor (Eds.), *Neuroergonomics: Cognitive neuroscience approaches to human factors and ergonomics*: Palgrave Macmillan.

Introduction

Chapter 1

Introduction

This dissertation was partly written using an application called Concentrate[1]. Concentrate is not a normal program, it does not let you write, draw, or send emails. No, it actually does not *do* anything. What it does is the opposite: it prevents you from doing too many things at the same time, it prevents you from too much multitasking. While it is often said that our "modern world is a multitasking world" (e.g., Salvucci, Taatgen, & Borst, 2009, p. 1), it has become more and more clear that multitasking is not necessarily a good thing. Many studies have shown that while we spend much of our time performing multiple tasks at the same time (e.g., Carrier, Cheever, Rosen, Benitez, & Chang, 2009; González & Mark, 2004), in general this leads to a decrease in performance (e.g., Alm & Nilsson, 1994; Brookhuis, de Vries, & de Waard, 1991; Gillie & Broadbent, 1989; Monk, Trafton, & Boehm-Davis, 2008). In fact, my promotor once said that one of the things that could explain the success of great scientists is their ability to *mono*task, their ability to concentrate on one thing at a time. That is where the aptly named program Concentrate comes into this thesis: it forced me to monotask, and thereby enabled me to finish this dissertation.

The topic of this dissertation is human multitasking, and in particular the so-called *problem state bottleneck*: one of the reasons why multitasking is often counter-productive. While in general humans are extremely good at multitasking – when do we truly do one task at a time? – in certain situations our ability to multitask breaks down. On the one hand, multitasking is obviously limited by physical constraints. We simply cannot look at two things at the same time, as texting cyclists prove daily during my ride to work. More interestingly, there are also limitations in our cognitive system that hinder multitasking. One of these is the problem state bottleneck: a limitation in processing intermediate representations that are necessary for a task. To give an everyday example, imagine you go to the living room to pick up John Anderson's latest book. You store the intermediate representation for this task – pick up Anderson's book – in your *problem state resource*, and walk to the living room. On the way, a friend calls you to ask where you are going to have dinner tonight. After finishing the call you find yourself standing in the living room, without the faintest notion about what you were going to pick up there. According to the theory that I will present in this dissertation, this is caused by a limitation in processing intermediate representations in our brain: the problem state bottleneck.

To support the idea of a problem state bottleneck, I will present several experiments and a computational theory of how intermediate representations are processed in our minds. As support for this theory, I will not only look at behavioral data, but also relate the theory to neuroimaging data. However, before turning to the problem state bottleneck, I will first give a short overview of existing multitasking theories. Multitasking has been investigated for over a century, and especially the threaded cognition theory (Salvucci & Taatgen, 2008, 2011) is of great importance for the current work. This theory will therefore be discussed in some detail below. Furthermore, both threaded cognition and

[1] *http://getconcentrating.com/*

the models in this dissertation were implemented in the cognitive architecture ACT-R (e.g., Anderson, 2007). I will therefore also briefly introduce cognitive architectures and ACT-R, followed by a discussion of how cognitive architectures can be combined with neuroimaging research. I will end this introduction with an overview of the other chapters.

Multitasking Theories

As early as 1931, Telford investigated interference due to human multitasking. He introduced the psychological refractory period (PRP) paradigm, and showed that people are slower to respond to the second of two tasks when these tasks have to be performed concurrently. Since Telford, many theories have been put forward to explain interference effects in multitasking (see for overviews, Meyer & Kieras, 1997a; Salvucci & Taatgen, 2008). Theories on multitasking can be divided into three general groups: bottleneck theories, resource theories, and cognitive control theories. Bottleneck theories assume fixed bottlenecks in human cognition that can only process one task at a time, causing interference when used by multiple tasks concurrently (e.g., Broadbent, 1958; Keele, 1973; Pashler, 1994; Welford, 1952). Theorists have identified several different bottlenecks, ranging from perceptual bottlenecks (e.g., Broadbent, 1958), to response-selection bottlenecks (e.g., Pashler, 1984; 1994), to motor bottlenecks (e.g., Keele, 1973). To unify these different bottleneck accounts, resource theories were introduced. These theories assume that attention can be flexibly employed, and that multitasking interference occurs when cognitive resources are required by multiple tasks at the same time, but not when tasks require different resources (e.g., Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1984, 2002). A third research tradition focuses on executive processing and cognitive control to explain multitasking interference (e.g., Baddeley, 1986; Cooper & Shallice, 2000; Meyer & Kieras, 1997a, 1997b; Norman & Shallice, 1986). In these theories, multitasking interference arises because of scheduling problems between tasks. That is, while tasks could in principle be carried out concurrently, executive control mechanisms enforce a certain task order, leading to interference. Using a cognitively bounded rational analysis, Howes, Lewis, and Vera (2009) have recently shown that to best account for at least the classical PRP effect (Schumacher et al., 1999; Telford, 1931) a theory needs cognitive control mechanisms, a motor bottleneck, and a response-selection bottleneck.

Based on the large body of data collected since the 1930s, detailed computational cognitive models of multitasking have been developed, ranging from concurrent multitasking (e.g., Kieras, Meyer, Ballas, & Lauber, 2000; Salvucci, 2005) to task switching (e.g., Altmann & Gray, 2008; Gilbert & Shallice, 2002; Monsell, 2003; Sohn & Anderson, 2001) to sequential multitasking (e.g., Altmann & Trafton, 2007; Salvucci, Monk, & Trafton, 2009). These computational models make it possible to predict the amount of interference between tasks on a quantitative level. Threaded cognition, a recent theory of human multitasking, combines all elements above and was implemented as a computational model (Salvucci & Taatgen, 2008, 2011). In combination with the cognitive architecture ACT-R (Anderson, 2007), it made a specific prediction that is the basis of this dissertation: it predicted the problem state bottleneck.

Threaded Cognition's Prediction

Threaded cognition is a general theory of human multitasking (Salvucci & Taatgen, 2008, 2011; Salvucci, Taatgen, et al., 2009). It assumes multiple different bottlenecks, and states that while multiple tasks can be performed concurrently, every resource in human cognition can only process one task at a time and therefore acts as a bottleneck when required by multiple tasks concurrently (cf. Byrne & Anderson, 2001). Depending on the requirements of the tasks at hand, these bottlenecks lead to different patterns of interference. Thus, the key assumption of threaded cognition is that although *several tasks* can be active at the same time, a particular *resource* can only be used by a *single task* at a time.

For instance, if two tasks want to use the visual system at the same time, only one of them can proceed and the other task will have to wait. In the case of the visual system this is quite obvious: we can only look at one object at a time. However, the same mechanism is assumed to hold for more central resources such as memory. According to threaded cognition, if two tasks want to retrieve a fact from memory at the same time, only one of them can proceed and the other task will have to wait. On the other hand, no interference is predicted if one task uses the visual system while another task retrieves a fact from memory. Thus, as long as the resource requirements of the different tasks do not overlap in time, threaded cognition predicts no interference, but as soon as a particular resource is concurrently needed by two or more tasks, that resource will act as a bottleneck and delay the execution of the combined process. This aligns with the intuition that if two tasks require the same cognitive constructs, the tasks will interfere (e.g., talking and reading both require our language faculties, while talking and walking require different resources).

It was shown that threaded cognition could account for interference caused by two peripheral bottlenecks (vision, motor) and two cognitive bottlenecks (procedural and declarative memory; Salvucci & Taatgen, 2008). In addition, based on its integration in the cognitive architecture ACT-R (e.g., Anderson, 2007), one more source of multitasking interference was predicted: the so-called problem state resource[2]. The problem state resource is used to maintain intermediate representations that are necessary for performing a task. For example, when calculating '2 + 3 × 4' mentally, one might use the intermediate representation '2 + 12'. According to the ACT-R theory, only a single intermediate representation can be maintained at a time, which should lead to interference when multiple representations are required concurrently.

Previously, we have presented results (Borst & Taatgen, 2007) that illustrated the potential role of the problem state resource as a bottleneck in multitasking. In that study, participants had to enter an address in a simulated navigation device while driving a simulated car. Both tasks had two versions: one that required maintaining

[2] *The* imaginal buffer *in ACT-R terminology.*

intermediate representations, and one in which there were no intermediate results. When both tasks required an intermediate representation, performance was slower and more error-prone than could be explained by the difficulty of the separate tasks alone, indicating a bottleneck in processing intermediate representations. However, the setup of that study was relatively under-constrained, making it difficult to derive precise conclusions. In this dissertation I will build on these results, and develop more precise experiments to investigate the problem state bottleneck. To account for the results of these experiments, I will present cognitive computational models that were implemented in the cognitive architecture ACT-R, and validate those models using neuroimaging data. I will now give a brief overview of these methodologies.

Methodologies

Cognitive Modeling & Cognitive Architectures

In 1973, Allen Newell boldly argued that psychology focuses too much on isolated tasks, and as a result does not progress much beyond solving 'small questions': He was worried that psychology would never integrate the results of the many separate experiments into a unified theory of human cognition (Newell, 1973). As a solution, he proposed cognitive architectures: unified theories of cognition in which computational processing models can be developed for a wide variety of tasks. The use of computational models forces one to specify theories at a very precise level, while developing different models within one theory ensures that models do not explain isolated phenomena, but that basic mechanisms are shared between tasks. Currently, there are several cognitive architectures in development, for instance Newell's SOAR (Newell, 1990), EPIC (Meyer & Kieras, 1997a), and ACT-R (Anderson, 2007).

Following Newell's suggestion, in this dissertation we implemented all models in the cognitive architecture ACT-R (Anderson, 2007). This means that the underlying resources of the presented models were validated previously, and could now be used to investigate our theory of the problem state bottleneck (e.g., Cooper, 2007; Newell, 1990). Moreover, because the architecture specifies how humans move the mouse or retrieve a fact from memory, we could develop models of complete tasks (in contrast to single mechanisms), enabling a direct comparison between human and model data. This is especially important for models of multitasking behavior, in which the interaction between cognitive and peripheral resources often causes the observed behavior (Kieras & Meyer, 1997; Van Maanen, Van Rijn, & Borst, 2009). In the next section we will introduce ACT-R, and explain how it maps onto the multiple resource theory of threaded cognition.

ACT-R

Figure 1.1 shows an overview of the ACT-R architecture (Anderson, 2005, 2007; Anderson, Bothell, et al., 2004). ACT-R assumes that the human cognitive system can be described as a system of largely independent modules (cognitive resources)

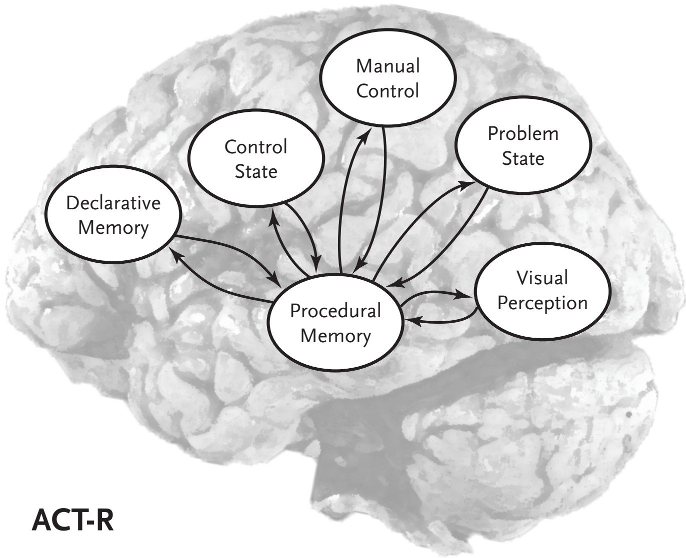


Figure 1.1 Core modules of the ACT-R cognitive architecture.

that interact through a central production system. It can perceive outside information through its visual module and its aural module (not shown in Figure 1.1), and act in the world through its manual module, which operates 'the hands' of ACT-R. To store information it uses a declarative memory store and a procedural memory store, while the control state maintains the current goal of the model. The problem state resource is used to store intermediate representations of a task, and is the main interest of this dissertation. According to our theory, it can only maintain at most one representation at a time, and therefore acts as a bottleneck in multitasking. In general, threaded cognition assumes that every ACT-R module constitutes a bottleneck: all modules can only proceed in a serial fashion, and therefore cause multitasking interference when required by multiple tasks concurrently (Byrne & Anderson, 2001; Salvucci & Taatgen, 2008, 2011).

Note that Figure 1.1 only shows the core modules of the architecture (the aural and vocal module are not shown). Other (or alternative) modules have also been developed, for instance to account for timing (Taatgen, Van Rijn, & Anderson, 2007; Van Rijn & Taatgen, 2008), blending in declarative memory (Lebiere, Gonzalez, & Martin, 2007), and robot perception (Trafton, Bugajska, Fransen, & Ratwani, 2008). Furthermore, while the ACT-R architecture mainly functions on a relatively high level (Newell's cognitive and rational bands, Newell, 1990; see also Anderson, 2002), of many modules more detailed lower-level versions have been proposed, for instance for declarative memory (Van Maanen, Van Rijn, & Taatgen, in press), procedural memory (Stocco, Lebiere, & Anderson, 2010), and visual perception (O'Reilly & Munakata, 2000; Salvucci, 2001).

Model-Based Neuroimaging

For a long time, the field of information processing psychology that argued for cognitive architectures essentially ignored the brain (Anderson, 2007). For instance, Newell states in 1980 that "symbolic behavior (and essentially rational behavior) becomes relatively independent of the underlying technology. Applied to the human organism, this produces a physical basis for the apparent irrelevance of the neural level to intelligent behavior." (Newell, 1980, p. 175). However, since the 1990s cognitive psychologists recognize the importance of the system in which intelligence is realized, and started connecting cognitive architectures to neuroimaging data. Anderson made this very explicit in his definition of a cognitive architecture in 2007 (p. 7): "A *cognitive architecture* is a specification of the structure of the brain at a level of abstraction that explains how it achieves the function of the mind."

One of the reasons for connecting cognitive architectures to neuroimaging data is that many models have a complexity that cannot be fully justified on the basis of behavioral measurements alone (e.g., Myung, 2000; Pitt & Myung, 2002; Roberts & Pashler, 2000). That is, there are so many degrees of freedom in developing a model that models are often under-constrained by behavioral data. To strengthen the constraints on cognitive models that are developed in the cognitive architecture ACT-R, a methodology was developed for mapping model activity on brain activity (for a concise explanation, see Anderson, Fincham, Qin, & Stocco, 2008). This way, models are not only constrained by behavioral data, but also by neuroimaging data. Figure 1.1 roughly shows the mapping between ACT-R's modules and the brain (for details, see Chapter 4 and 5).

Usually, the connection between ACT-R's modules and fMRI data is implemented using predefined Regions-Of-Interest, which provide a mapping between the brain and components of the architecture (Anderson, 2007; Anderson et al., 2008). For instance, activity in the motor resource of ACT-R should correspond to neural activity in a predefined region in the motor cortex (see Figure 1.1). We applied this methodology in Chapter 4 of this dissertation to investigate whether our model made plausible fMRI predictions.

More recently, a new methodology has emerged in fMRI research: model-based fMRI (e.g., Gläscher & O'Doherty, 2010; O'Doherty, Hampton, & Kim, 2007). This analysis technique shows regions in the brain where neural activity significantly correlates with model activity. In Chapter 5, we applied this technique for the first time to a model developed in a cognitive architecture. Model-based fMRI is a promising new method, because it gives a functional explanation of fMRI data by directly linking the data to model constructs (which naturally perform required functions of the model). Functional neuroimaging, especially fMRI, has often been criticized of not contributing anything significant to our understanding of the mind as there is no direct mapping between data and function (e.g., Coltheart, 2004; Coltheart, 2006; Fodor, 1999; Harley, 2004; Page, 2006; but see e.g., Aue, Lavelle, & Cacioppo, 2009; Friston, 2009; Hagoort, 2008; Henson, 2005, 2006; Jonides, Nee, & Berman, 2006; Logothetis, 2008). As Fodor put it forcefully: "If the mind happens in space at all, it happens somewhere north of the neck", on which basis he discounted most fMRI research from explaining anything about our cognitive system (Fodor, 1999). However, especially by combining model-based fMRI with models grounded in a cognitive architecture (which we have shown is possible in Chapter 5), we can at least partly avoid these criticisms, and use fMRI to learn more about the functioning of our cognitive system.

Overview of this Dissertation

As stated above, this dissertation is about the problem state bottleneck. I will present behavioral, model-based, and neuroimaging support for the existence of this bottleneck. First, I will present three behavioral experiments and accompanying cognitive models in Chapter 2. These experiments provide initial support of a problem state bottleneck. In Chapter 3, the behavioral support is extended with an experiment that shows how the problem state bottleneck can be bypassed. In addition, pupil dilation data will be presented in Chapter 3, to show that the bottleneck is associated with an increase in mental workload. In Chapter 4 and 5, I turn to neuroimaging data to validate the cognitive model that was presented in Chapter 1. First, in Chapter 4, a region-of-interest analysis is applied to test *a priori* predictions of the cognitive model. Second, in Chapter 5, the novel model-based fMRI analysis technique is used to show where in the brain the different resources of the model are most likely represented. To conclude, in Chapter 6, our final theory of how intermediate representations are processed in the mind will be presented. This theory will be backed up with data presented in the other chapters of this thesis, and with data of two new behavioral experiments.

The Problem State: A Cognitive Bottleneck in Multitasking

In which we present first support for a single-sized problem state resource in the form of three behavioral experiments.

This chapter was previously published as:

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.

2

Chapter

2 Chapter

The Problem State: A Cognitive Bottleneck in Multitasking

Abstract

The main challenge for theories of multitasking is to predict when and how tasks interfere. Here we focus on interference related to the problem state, a directly accessible intermediate representation of the current state of a task. On the basis of Salvucci and Taatgen's (2008) threaded cognition theory, we predict interference if two or more tasks require a problem state, but not when only one task requires one. This prediction was tested in a series of three experiments. In Experiment 1, a subtraction and text-entry task had to be carried out concurrently. Both tasks were presented in two versions: one that required maintaining a problem state and one that did not. A significant over-additive interaction effect was observed, showing that the interference between tasks was maximal when both tasks required a problem state. The other two experiments tested whether the interference was indeed due to a problem state bottleneck, instead of cognitive load (Experiment 2; an alternative subtraction and text entry experiment) or a phonological loop bottleneck (Experiment 3; a triple-task experiment that added phonological processing). Both experiments supported the problem state hypothesis. To account for the observed behavior, computational cognitive models were developed using threaded cognition within the context of the cognitive architecture ACT-R (Anderson, 2007). The models confirm that a problem state bottleneck can explain the observed interference.

Introduction

Some tasks can be performed together effortlessly, like walking and talking, while other tasks interfere with each other, like car driving and phoning, while again other combinations of tasks are nearly impossible to do concurrently, like writing a manuscript and talking to a colleague. Intuitively, it seems clear why some tasks interfere with each other and some do not: the more overlap in cognitive constructs between tasks, the more interference. For instance, writing a paper and talking to a colleague both use language faculties, resulting in major interference between the tasks.

Psychologists have been formally investigating multitasking behavior at least since the 1930s (e.g., Telford, 1931; see Meyer & Kieras, 1997a for an excellent review). Based on the large body of research collected since the 1930s, detailed cognitive models of multitasking have been developed, ranging from concurrent multitasking (e.g., Kieras et al., 2000; Salvucci, 2005) to task switching (e.g., Altmann & Gray, 2008; Gilbert & Shallice, 2002; Sohn & Anderson, 2001; see also Monsell, 2003) to sequential multitasking (e.g., Altmann & Trafton, 2007). These computational models make it possible to predict the amount of interference between tasks on a quantitative level. To unify several areas of multitasking, Salvucci and Taatgen (2008) recently proposed a new theory of multitasking behavior, threaded cognition, which accounts for concurrent multitasking as well as for sequential multitasking (see also Salvucci, Taatgen, et al., 2009). Threaded cognition was implemented in the cognitive architecture ACT-R (Anderson, 2007), enabling researchers to make formal models of multitasking behavior.

In threaded cognition, tasks can use several distinct cognitive resources, such as vision, manual operations, or memory. These resources can operate in parallel, but are themselves serial in nature (cf. ACT-R, Anderson, 2007; Byrne & Anderson, 2001). Because of this seriality, a resource can only be involved in one operation at a time, but multiple resources can be active at the same time. This within-resource seriality but between-resource parallelism holds regardless of whether the resources are recruited for a single task (e.g., physically moving a disc in a Towers of Hanoi problem while at the same time using memory to plan the next move) or whether the resources are recruited for different tasks (manually tuning the car-audio system while at the same time visually processing the road in front of the car). Thus, the key assumption related to multitasking in threaded cognition is that although several tasks can be active at the same time, a particular resource can only be used by a single task at a time. For instance, if two tasks want to use the visual system at the same time, only one of them can proceed and the other task will have to wait. In the case of the visual system this is quite obvious: we can only look at one object at a time. However, the same mechanism is assumed to hold for more central resources, like memory. For example, if two tasks want to retrieve a fact from memory at the same time, only one task can proceed; the other task will have to wait. On the other hand, no interference is predicted if one task wants to use the visual system, and one task wants to retrieve a fact from memory. Thus, as long as the resource requirements of the different tasks do not overlap in time, threaded cognition predicts no interference, but as soon as a particular resource is concurrently needed by two or more tasks, that resource will act as a bottleneck and delay the execution of the combined process. This aligns with the intuition that if two tasks require the same cognitive constructs, the tasks will interfere.

Salvucci and Taatgen (2008) discussed two peripheral bottlenecks (the visual and motor system) and two central cognitive bottlenecks (declarative and procedural memory; cf. "attentional limitations", Pashler & Johnston, 1998). In this article, we discuss a third central cognitive resource that can result in significant interference, both in terms of decreased speed and increased errors: the problem state. The problem state resource is used to maintain intermediate mental representations that are necessary for performing a task. For instance, while solving an algebra problem like $2x - 5 = 8$, the problem state can be used to store the intermediate solution $2x = 13$. The problem state resource is assumed to be limited to only one coherent 'chunk' of information (Anderson, 2005, 2007), and will therefore cause interference when multiple tasks concurrently require its use. However, not all tasks require the use of a problem state. If no intermediate results need to be stored (e.g., solving one step of the algebra problem $2x = 8$ immediately results in the required answer) or all necessary information is present in the world (e.g., if the intermediate steps can be selected from and are displayed on a computer screen), there is no need for maintaining a mental representation.

Previously, we have presented results (Borst & Taatgen, 2007) that illustrated the potential role of the problem state resource as a bottleneck in multitasking. In this study, participants had to type in an address in a simulated navigation device while driving a simulated car. The task required switching back and forth between driving and operating the navigation device. Both tasks had two versions: one that required maintaining intermediate results, and one in which there were no intermediate results. More specifically, in the driving task participants had to memorize the turns to take at the next intersections in one condition, while in the other condition arrows pointed out the route. In the navigation task, the two conditions differed in whether the participants had to memorize the full address before entering it, or whether the device would show what letter to press next. When both difficult conditions were combined, performance was much slower and more error-prone than could be explained by the difficulty of the separate tasks alone. That study suggested that combining certain tasks yield additional costs in terms of time and errors. However, the setup of the study was relatively under-constrained, making it difficult to derive precise conclusions.

In the current article, we investigate whether the problem state resource constitutes a bottleneck in a more constrained setting. In the first experiment, participants performed a complex dual-task. Data of this experiment are in line with predictions derived from a problem state bottleneck–based theory. However, to test whether the results of Experiment 1 were caused by cognitive load effects (e.g., Logan, 1979), Experiment 2 controls for cognitive load over the different conditions, while in Experiment 3 another possible explanation involving the phonological loop was investigated. Experiments 2 and 3 both provide corroborating support for a problem state bottleneck account. The experimental findings are supported by computational cognitive models, which show that a problem state bottleneck can explain the observed interference effects. Before we describe the experiments, we will introduce the problem state resource and the threaded cognition theory in more detail.

The Problem State Resource

In our terminology, the problem state resource is used for storing intermediate information that is necessary for performing a task. Information in the problem state resource is directly accessible for the task at hand, while it takes time to retrieve facts from declarative memory (cf. ACT-R, Anderson, 2007). For instance, while mentally solving an algebra problem like $3x - 12 = 0$, the problem state can be used to store the intermediate solution $3x = 12$; and when asking for directions, the problem state can be used to store at which street you should turn to arrive at your destination. If this information is present in the world, that is, if you work out an algebra problem on paper or follow road signs to a destination, it is not necessary to maintain a problem state.

The concept of the problem state stems from a series of neuroimaging experiments by Anderson and colleagues, who found BOLD activity in the posterior parietal cortex that correlates with the transformation of mental representations (e.g., Anderson, 2005; Anderson, Albert, & Fincham, 2005; Anderson, Qin, Sohn, Stenger, & Carter, 2003; Sohn et al., 2005). They concluded on this basis that a separate resource exists for maintaining and transforming mental representations.

The problem state construct is closely linked to mental states as used by Altmann et al. in their cognitive control model and memory for goals theory to explain task switching and task interruption behavior (Altmann & Gray, 2008; Altmann & Trafton, 2002, 2007). However, where in their case mental representations constitute both the goal and the problem state of a task, in threaded cognition (also in the current version of ACT-R, e.g., Anderson, 2005, 2007) these mental representations have been split into a goal state that only maintains the state of the current goal, and a problem state that maintains temporary intermediate information necessary for doing the task (but see Salvucci, Taatgen, et al., 2009, about how these theories can be reconciled). The problem state is also related to the 'episodic buffer' in Baddeley's (2000) extension of the classical working memory model of Baddeley and Hitch (1974). This buffer serves the function of a "limited capacity temporary storage system that is capable of integrating information from a variety of sources" (*p.* 421), which was previously part of the 'central executive' (Baddeley, 2003). This construct is very similar to ACT-R's problem state resource, in the sense that both systems can integrate information from different sources (perceptual and long-term memory) and temporarily store the outcome for further processing.

The Threaded Cognition Theory

Threaded cognition (Salvucci & Taatgen, 2008) is an integrated theory of human multitasking. In threaded cognition, every task is represented by a so-called cognitive thread. For instance, in the case of driving a car and operating a navigation device, one thread would represent steering the car and another thread would represent operating the navigation device. A thread is associated with the goal of a task, which serves as a key to mobilize associated task knowledge (e.g., declarative and procedural memory that is necessary for performing the task). Although multiple threads can be active at a time, only a single procedural processor is available; thus, although multiple threads are active in parallel, only one thread can use the procedural processor at a time (compare this to multiple programs running on a single CPU on a computer: while the CPU can only process one instruction at a time, programs act as if they were executed concurrently). Furthermore, if a thread needs to use a cognitive resource such as vision or memory, it can only be selected for execution if that resource is available. Thus, while the threads act in parallel and are not governed by any supervisory executive control structure, they are constrained by the available resources. (For a similar approach, but from a more mathematical point of view, see Liu, Feyen, & Tsimhoni, 2006.)

The threaded cognition theory is implemented in the cognitive architecture ACT-R (Anderson, 2007). ACT-R describes human cognition as a set of independent modules that interact through a central production system. For instance, it uses visual and aural modules for perception and a motor module to interact with the world. Besides these peripheral modules, ACT-R also has a number of central cognitive modules: the procedural module that implements the central production system, the declarative memory module, the goal module, the timing module (Taatgen et al., 2007; Van Rijn & Taatgen, 2008) and the problem state module[1]. All modules operate in parallel, but each module in itself can only proceed serially (Byrne & Anderson, 2001). Thus, the visual module can only perceive one object at a time and the memory module can only retrieve one fact at a time.

A task is represented in ACT-R by the contents of the goal module and the problem state module (Anderson, 2007). In the case of solving an algebra problem like $8x - 5 = 7$, the goal module can hold for instance 'algebra - unwinding', while the problem state module can be used to hold the intermediate solution $8x = 12$. Thus, the goal module holds the current state of a task, while the problem state module holds intermediate information necessary for performing the task. In line with the serial processing in the other modules, the goal module can only hold a single goal and the problem state module can only hold a single problem state at a time.

Threaded cognition extends ACT-R by allowing for multiple parallel goals, and thus multiple tasks (threads), to be active. This translates into the assumption that the goal module in ACT-R can represent several goals at the same time. However, the other modules can still only do one thing at a time, which means that they can only be used by one thread at a time. The modules are shared on a first-come-first-served basis: a thread will 'greedily' use a module when it needs it, but also will let go of it 'politely', that is, as soon as it is done with it. The seriality of the modules results in multiple potential bottlenecks: when two threads need a module concurrently, one thread will have to wait for the other.

[1] *Sometimes referred to as 'imaginal module' or 'problem representation module'.*

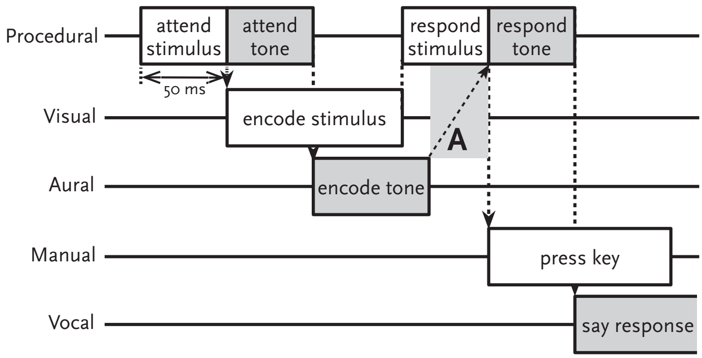


Figure 2.1 Example processing stream in threaded cognition. White boxes depict a visual–manual task, grey boxes an auditory–vocal task. The 'A' represents interference, caused by both threads needing the procedural resource at the same time.

In Figure 2.1 an example processing stream of a dual-task in threaded cognition is shown: white boxes depict a task in which a key-press is required in response to a visual stimulus, and grey boxes depict a task in which a vocal response is required in response to an auditory stimulus. The *x*-axis represents time and boxes represent the period of time during which a resource is used. Both tasks start by activating production rules to initiate attending the respective stimuli, after which the encoding process starts in both the visual and the aural module. The grey area marked A indicates interference, caused by the concurrent request for the procedural module after the respective encoding steps. As the visual–manual task already uses the procedural module, the auditory–vocal task has to wait. Thus, if multiple threads require a resource at the same time, interference is observed.

Salvucci and Taatgen (2008) presented cognitive models that account well for dual-tasking in a number of different domains, ranging from simple Psychological Refractory Period (PRP) tasks to driving a car and using a cell phone concurrently. These models showed that bottlenecks in perceptual and motor resources in addition to bottlenecks in two more central cognitive resources (procedural and declarative memory) account for a wide range of multitasking interference phenomena (but see for a more detailed account of interference in the motor system e.g., Albert, Weigelt, Hazeltine, & Ivry, 2007; Diedrichsen, Hazeltine, Kennerley, & Ivry, 2001). Although multiple bottlenecks are identified, not all bottlenecks result in the same interference profiles. The severity of the interference depends on the particular resource: procedural memory is very fast and therefore only leads to delays in the order of 50 ms (but see Taatgen, Juvina, Schipper, Borst, & Martens, 2009, for an example where interference caused by the procedural resource explains counter-intuive results in an attentional blink dual-task). On the other hand, interference due to declarative memory and the visual and motor system leads to pronounced decreases in speed in the order of 200–500 ms.

Salvucci and Taatgen (2008) did not investigate the role of the problem state resource in multitasking. Because many tasks require the maintenance of intermediate representations and because this maintenance is required for relatively long periods of time (i.e., several seconds), we hypothesize that the problem state is a important source of interference in multitasking. We will now turn to three experiments that test this hypothesis.

Experiment 1: Subtraction & Text-Entry

In Experiment 1, participants had to perform two tasks concurrently: a subtraction task and a text-entry task. Both tasks were presented in two versions: an easy version in which there was no need to maintain a problem state, and a hard version where participants had to maintain a problem state from one response to the next. Thus, the experiment has a 2 × 2 factorial design (Subtraction Difficulty × Text-Entry Difficulty). As threaded cognition claims that the problem state resource can only be used by one task concurrently, we hypothesized that when a problem state is required in both tasks (the *hard–hard* condition), participants will be significantly slower or make more errors than in the other conditions. On the other hand, if just a single task requires a problem state, no interference is to be expected on behalf of the problem state. Thus, we expected an over-additive interaction effect of task difficulty.

Method

Participants

Fifteen students of the University of Groningen participated in the experiment for course credit (10 female, age range 18–31, mean age 20.1). All participants had normal or corrected-to-normal visual acuity. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

Design

During the experiment, participants had to perform a subtraction task and a text-entry task concurrently. The subtraction task was shown on the left side of the screen, the text-entry task on the right (see Figure 2.2). Participants had to alternate between the two tasks: after a digit, the subtraction interface was disabled, forcing the participant to subsequently enter a letter. After entering a letter, the text-entry interface was disabled and the subtraction interface became available again.

The subtraction task is shown on the left side of Figure 2.2. Participants had to solve 10-column subtraction problems in standard right to left order; they had to enter the digits with their left hand using the keyboard. In the easy, no problem state version, the upper term was always larger or equal to the lower term; these problems could be solved without borrowing. In contrast, the hard version (as shown in Figure 2.2) required participants to borrow six times. The assumption is that participants use their problem state resource to keep track of whether a borrowing is in progress.

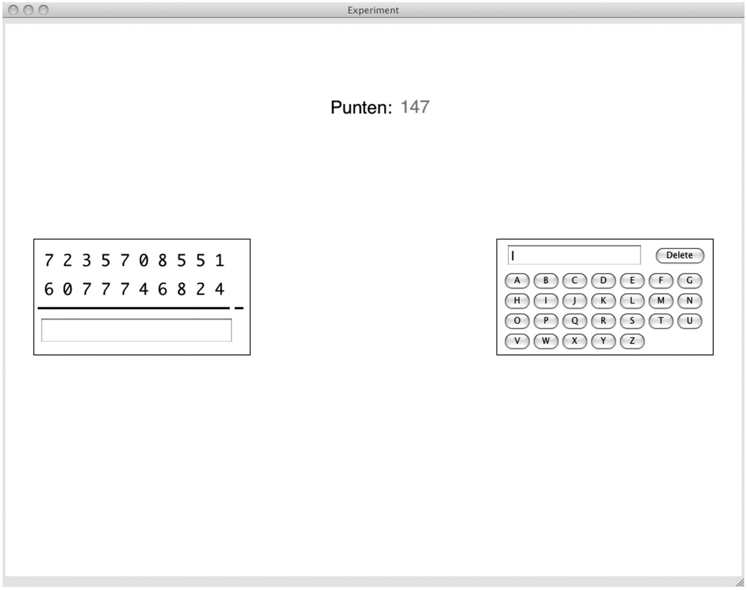


Figure 2.2 Screenshot of Experiment 1.

The second task in the experiment is text-entry. The interface is shown on the right in Figure 2.2: by clicking on the on-screen keypad 10-letter strings had to be entered. In the easy version of the text-entry task, the strings were presented one letter at a time. Participants saw one letter appear on the screen (for example the 'I' in Figure 2.2) and had to click the corresponding button on the keypad. As soon as a button was pressed, the text-entry keypad was disabled and the mouse pointer was hidden to prevent participants from putting the pointer on the next letter. When the text-entry task was re-enabled, the mouse pointer appeared again in the location where it had been hidden.[2] Participants could only enter the next letter after the next subtraction column was responded to. After 10 letters had been entered, the trial ended automatically. In the hard version, a 10-letter word appeared at the start of a trial. When the participant clicked on the first letter, the word disappeared and had to be entered without feedback (thus, participants could neither see what word they were entering, nor what they had entered, the text-entry screen remained blank until the end of the trial). Otherwise, both conditions were identical. In the hard version, we assume that participants need their problem state resource to keep track of what word they were entering and at which position they are (e.g., "*informatie*, 4th position").

As is shown in Figure 2.2, participants could earn points (*punten* in Dutch). Participants started out with 200 points. While performing the tasks, the counter at the top of the screen decreased by 2 points per second. For every correct letter or digit 10 points were added to the total (addition was done after finishing the complete trial).

---

[2] *Participants could have used the mouse to indicate what the last letter was that they entered. However, that would have made it harder to find our results, as that means that they would have maintained less information mentally (only the word, not the position within the word).*

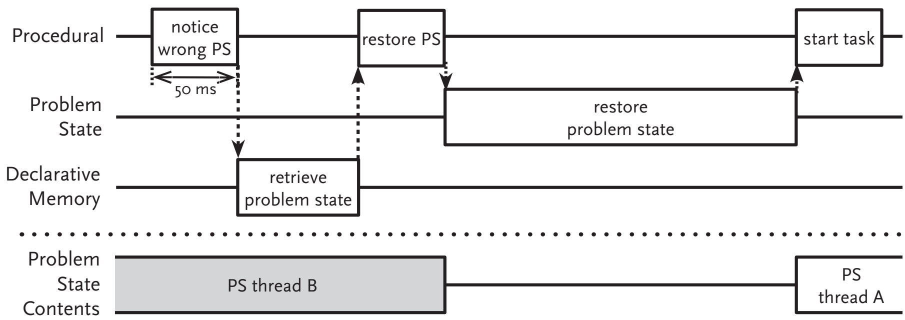


Figure 2.3 Processing stream of replacing a problem state. PS = problem state.

At the end of a trial a feedback display was shown to the participants, indicating how many points they gained per task in the current trial. In effect, to score a high amount of points participants had to act both quickly and accurately.

Stimuli and Apparatus

The stimuli for the subtraction task were generated anew for each participant. The subtraction problems in the hard version always featured six borrowings, and resulted in 10-digit answers. The 10 letter words for the hard version of the text-entry task were handpicked from a list of high frequent Dutch words (CELEX database, Baayen, Piepenbrock, & Van Rijn, 1993), to ensure that similarities between words were kept to a minimum. These stimuli were also used in the easy text-entry task, except that the letters within the words were scrambled (under the constraint that a letter never appeared twice in a row). Thus, participants entered random sequences of letters. This did not introduce difficulties, because the participants never saw the complete letter-sequences but had to enter the letters one-by-one. By scrambling the words, we controlled for letter-based effects, while preventing the use of alternative strategies to predict the next letter.

The experiment was presented full screen on a 19-inch monitor. The width of both the subtraction interface and the text-entry interface measured 9 centimeters, while the space between the two tasks was 10 cm; the height of the interfaces was 4.8 cm (see also Figure 2.2). Participants were sitting at a normal viewing distance, about 75 cm from the screen.

Procedure

A trial started with the appearance of the two tasks. Participants could choose which task to start with; after the first response they were required to alternate between the tasks. After the last response of a task within a trial a feedback display appeared, showing how many letters or digits had been entered correctly. After giving the last response of a trial, there was a 5 second break until the next trial.

Before the experiment, participants completed 6 practice trials for the separate tasks, and 4 for the dual-task. The experiment consisted of three blocks. Each block consisted of four sets of three trials per condition. These condition-sets were randomized within a block, with the constraint that the first condition of a block was different from the last condition in the previous block. Thus, the participants had to perform 36 trials, presented semi-randomly. The complete experiment lasted approximately 45 minutes. Halfway the experiment participants could take a short break.

Model

We will first describe the computational cognitive model[3] that we developed for the task, after which the behavioral and modeling results will be presented side by side. The model was developed in the ACT-R cognitive architecture (Anderson, 2007; Anderson, Bothell, et al., 2004), using threaded cognition (Salvucci & Taatgen, 2008).

Of particular importance for the tasks at hand is ACT-R's problem state module. This module can hold a problem state, accessible at no time cost. However, changing a problem state takes 200 ms (Anderson, 2007). Because the problem state module can only hold one chunk of information, the module's contents have to be exchanged frequently when multiple tasks require a problem state. When the problem state is replaced, the previous problem state is automatically moved to declarative memory so that it can be restored when the other thread needs it. Figure 2.3 displays an example processing stream of problem state replacement. The white boxes represent Task A that requires the problem state resource, while the grey box represents the problem state of Task B, occupying the resource at the start of the example. First, the white task notes ("notice wrong PS") that the problem state resource does not contain its own associated problem state, and therefore initiates a process to retrieve this problem state from declarative memory. This retrieval takes a certain amount of time, after which a production rule ("restore PS") fires to start restoring the retrieved problem state to the problem state resource. This takes a fixed 200 ms. After this initialization process, the white task can start with its actual operation. The total time to replace the problem state resource is thus 200 ms plus the time for the retrieval plus 100 ms for the "notice wrong PS" and "restore PS" production rule executions. Thus, when multiple tasks need the problem state resource, the execution time of tasks is increased considerably per change of task. An additional effect of this exchange of problem states is that because problem states need to be retrieved from memory, it is possible that a task retrieves an older, and thus incorrect problem state from memory, resulting in behavioral errors.

The two tasks in the experiment were implemented as two threads: a subtraction thread and a text-entry thread. Both threads use the visual module to perceive the stimuli and the manual module to operate the mouse and the keyboard. In the easy condition of the subtraction task, the model perceives the digits, retrieves a fact from memory (e.g., 5 - 2 = 3) and enters the difference. In the hard condition, the general

[3] *Available for download at http://www.ai.rug.nl/~jpborst/models/.*

process is the same. However, if the model retrieves a fact from memory and notices that the outcome is negative (e.g., 3 - 6 = -3), the model will add 10 to the upper term, store in its problem state that a borrowing is in progress, and retrieve a new fact (13 - 6 = 7). When the model encounters a negative subtraction outcome for the first time in a trial, it notes in its goal state that it is performing the hard version of the task ("subtraction – hard"). This ensures that the model checks for the appropriate problem state at the start of each subsequent response-sequence (as the problem state indicates whether a borrowing is in progress). If a borrowing is in progress, the model first subtracts 1 from the upper term before the initial retrieval is made.

In the easy version of the text-entry task, the model perceives the letter and clicks on the corresponding button. In the hard version, the model has to know the target word and the current position within that word. Thus, it requires the problem state resource to store what word it is entering and at which position of the word it is ("*informatie*, 4th position"). If the model performs a trial in the hard condition, it will use the word and position in its problem state to come up with the next letter. To simulate the spelling processes required to come up with "letter 5 from the word *informatie*", we have assumed that an additional declarative retrieval is necessary that links the current position to the next letter. As spelling words is not the focus of this article, we did not model this in detail, but instead assumed an additional retrieval. After the model has determined the next letter, it clicks the appropriate button and updates its problem state to reflect that it is one position further in the word.

The ACT-R theory predicts the time it takes to perceive a stimulus, to press a key and to move the mouse, and to retrieve facts from declarative memory, which makes it meaningful to incorporate these parts of the task in the model. These elements of ACT-R have been tested and validated separately, many examples can be found at http://act-r.psy.cmu.edu/. Instead of discussing all details here, we refer the reader to Anderson (2007) for more information.

Because the model requires two problem states that need to be exchanged at each trial in the *hard–hard* condition, and either zero (*easy–easy*) or one (*easy–hard*, *hard–easy*) in the other conditions, it predicts an over-additive effect of task difficulty on response times. Possibly, the number of errors will also increase, depending on whether older and incorrect problem states are retrieved frequently.

Results

Only the data of the experimental phase were analyzed. Two participants did not adhere to task instructions and were removed from the dataset. Outliers in response times faster than 250 ms and slower than 9000 ms were removed from the data, after which we removed data exceeding 3 standard deviations from the mean per condition per participant (in total, 2.0% of the data was removed). All reported *F*- and *p*-values are from repeated-measure ANOVAs, all error bars depict standard errors, effects were judged significant if they reached a .05 significance level. Accuracy data were transformed using an arcsine transformation before performing ANOVAs. Figure 2.4 shows the main results, black bars depict experimental data, grey bars model data.

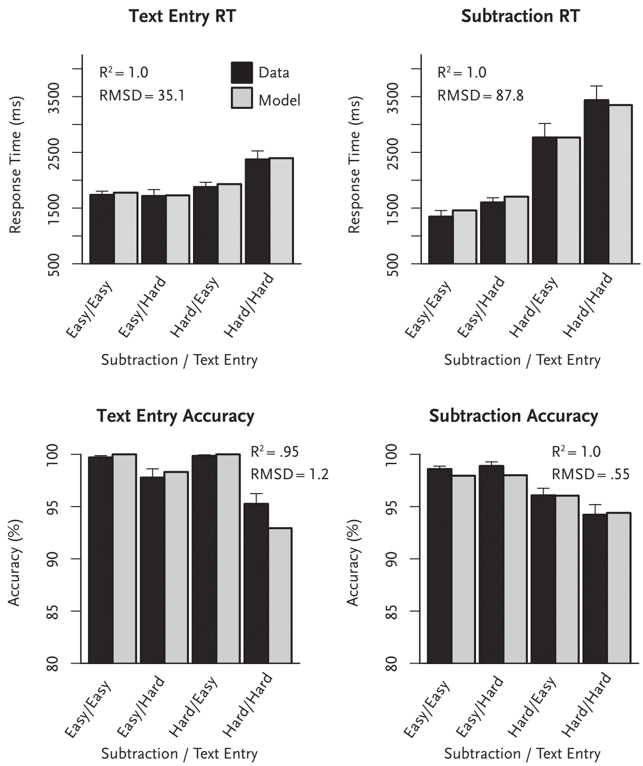


Figure 2.4 Results of Experiment 1. RMSD = root mean squared deviation.

Response Times

Response time on the text-entry task was defined as the time between entering a digit in the subtraction task and clicking on a button of the text-entry task. First responses of each trial were removed. The upper left panel of Figure 2.4 shows the results. First, an interaction effect between Subtraction Difficulty and Text-Entry Difficulty ($F(1,12) = 22.15$, $p < .001$, $\eta_p^2 = .65$) was found. Next, we performed a simple effects analysis, showing an effect of Text-Entry Difficulty when subtraction was hard ($F(1,12) = 10.78$, $p < .01$, $\eta_p^2 = .47$), and an effect of Subtraction Difficulty when text-entry was hard ($F(1,12) = 47.16$, $p < .001$, $\eta_p^2 = .80$). The other simple effects did not reach significance: Text-Entry Difficulty when subtraction was easy ($F(1,12) = 1.88$, $p = .20$, $\eta_p^2 = .14$) and Subtraction Difficulty when text-entry was easy ($F(1,12) = 3.35$, $p = .09$, $\eta_p^2 = .22$). Thus, there was an over-additive interaction effect of task difficulty on response times of the text-entry task; participants were slowest to respond in the *hard–hard* condition, no other effects were found.

Figure 2.4, upper right panel, shows the average response times on the subtraction task. This is the time between clicking a button in the text-entry task and entering a digit in the subtraction task. Again, first responses of a trial were removed, as were responses that occurred in the hard conditions before a borrowing had taken place, as those are in effect easy responses. An interaction effect between Subtraction Difficulty and Text-Entry Difficulty was observed ($F(1,12) = 6.24$, $p = .03$, $\eta_p^2 = .34$). A simple effects analysis revealed that all simple effects were significant: Subtraction Difficulty when text-entry was easy ($F(1,12) = 69.04$, $p < .001$, $\eta_p^2 = .85$), Subtraction Difficulty when text-entry was hard ($F(1,12) = 111.64$, $p < .001$, $\eta_p^2 = .90$), Text-Entry Difficulty when subtraction was easy ($F(1,12) = 11.65$, $p < .01$, $\eta_p^2 = .49$), and Text-Entry Difficulty when subtraction was hard ($F(1,12) = 11.81$, $p < .01$, $\eta_p^2 = .50$). Thus, the more difficult the tasks, the higher the response times, with an over-additive effect in the *hard–hard* condition, reflected by the interaction.

Accuracy

Figure 2.4, lower left panel, shows the accuracy on the text-entry task, in percentage correctly entered letters. Both main effects were significant: Subtraction Difficulty ($F(1,12) = 7.31$, $p = .02$, $\eta_p^2 = .38$) and Text-Entry Difficulty ($F(1,12) = 21.57$, $p < .001$, $\eta_p^2 = .64$). The interaction effect between Subtraction Difficulty and Text-Entry Difficulty shows a trend towards significance ($F(1,12) = 4.65$, $p = .052$, $\eta_p^2 = .28$). Thus, accuracy on the text-entry task decreased as a function of both Text-Entry Difficulty and Subtraction Difficulty, with a trend towards a stronger decrease when both tasks were hard.

In the lower right panel of Figure 2.4, the accuracy on the subtraction task is shown. Here, a significant interaction effect between Subtraction Difficulty and Text-Entry Difficulty was observed: $F(1,12) = 10.50$, $p < .01$, $\eta_p^2 = .47$. A simple effects analysis subsequently revealed that three simple effects reached significance: Text-Entry Difficulty when subtraction was hard ($F(1,12) = 6.68$, $p = .02$, $\eta_p^2 = .36$), Subtraction Difficulty when text-entry was easy ($F(1,12) = 7.17$, $p = .02$, $\eta_p^2 = .37$), and Subtraction Difficulty when text-entry was hard ($F(1,12) = 87.7$, $p < .001$, $\eta_p^2 = .88$). Text-Entry Difficulty when subtraction was easy did not reach significance ($F(1,12) = 3.64$, $p = .08$, $\eta_p^2 = .23$). Thus, when subtraction was hard accuracy was lower, but this effect was even stronger when text-entry was hard as well.

Model

The grey bars in Figure 2.4 show the results of the model. It resembles the empirical data closely (R^2- and Root Mean Squared Deviation-values are displayed in the graphs). The model shows the same interaction effects as the data, both in response times and accuracy. To fit the model, we estimated how long memory retrievals take[4] and how often incorrect memories are retrieved (i.e., retrieving problem states from declarative memory in the *hard–hard* condition, but also arithmetic errors like 9 - 6 resulting in 2 instead of 3). The incorrect retrievals were modeled in a similar fashion as in Anderson,

[4] *ACT-R's latency factor was set to .3 and activation noise to .1. Furthermore, subtraction facts were divided into two groups, one group of facts having a minuend under 10, and one group above 10. A third group was formed by the addition facts. The activation levels for those three groups of arithmetic facts were scaled to fit the participant group's behavior. The exact values of these parameters can be found in the model code online at http://www.ai.rug.nl/~jpborst/models/.*

Reder and Lebiere's (1996; see also Lebiere, 1999) model that accounts for arithmetic errors. All other parameters were kept at the default values of ACT-R 6.0 (Anderson, 2007; see also Anderson, Bothell, Lebiere, & Matessa, 1998).

As explained in detail above, the interaction effect in the model data is driven by the problem state bottleneck in the *hard–hard* condition. The model also accounts for the different reaction time patterns in the two tasks: in the subtraction task there is a large main effect of Subtraction Difficulty, while there is no such effect in the response times of the text-entry task. The model accounts for this by assuming that in the hard subtraction task, participants have to retrieve multiple facts from declarative memory to be able to enter a digit, as opposed to the easy subtraction task, in which only one fact has to be retrieved. In the text-entry task, on the other hand, there is no such difference between the easy and the hard task: in the easy version the model has to look at the display to see what letter it has to enter, while in the hard version it has to retrieve an order fact from memory and use information from its problem state to enter a letter. The timing of those processes is similar, resulting in the absence of a main effect of Text-Entry Difficulty on the response times in the text-entry task (cf. the upper left panel of Figure 2.4).

The model keeps track of the task condition in its goal state ("subtraction – hard", see the model description above). This state was set as soon as the thread noticed that it was performing a hard trial: initially it was always set to easy, but when the model came across a borrowing in the subtraction task or a complete word in the text-entry task, it would be set to hard. Did the participants also keep track of the task condition? We compared response times of subtraction columns from the hard condition in which no borrowing is in progress and in which no new borrowing is necessary (i.e., in every way comparable to columns in the easy condition, except that a borrowing has occurred more than one column back; for instance the left-most column of Figure 2.2), to columns of the easy subtraction condition. The difference in response time (2256.3 vs. 1466.3 ms) is significant (paired *t*-test, $t(12) = -10.10$, $p < .001$). This seems to indicate that participants were sensitive to the context of the current trial (i.e., the task condition): the task in these no-borrow columns in the hard subtraction conditions is exactly the same as in the easy subtraction task, only the context is different. This is consistent with the model's keeping-track account, which always checks whether a borrowing is in progress in the hard trials but not in the easy trials. For the model, this results in a difference in response times between hard responses that are comparable to the easy task and easy responses, although the difference is smaller (1762.4 vs. 1583.8 ms).

Discussion

The interaction effects in the data are in agreement with our model predictions: an over-additive effect of task difficulty on response times and error rates (a trend in the case of accuracy on the text-entry task). As described above, the model accounts for these interaction effects by proposing a problem state bottleneck that results in higher response times on the one hand (caused by constantly replacing the problem state) and higher error rates on the other (caused by retrieving older, incorrect problem states). The errors in the other conditions are caused by sometimes retrieving wrong facts from memory (i.e., 9 - 6 results in 2 instead of 3, see Anderson et al., 1996; and Lebiere, 1999).

Another interesting observation is the effect of condition of one task on the other task. More specifically, there is a significant effect of Text-Entry Difficulty on the reaction times of the subtraction task when subtraction was easy, and a marginal significant effect (p = .09) of Subtraction Difficulty on reaction times of the text-entry task when text-entry was easy. As can be seen in Figure 2.4, the model captures these effects. In the model, these effects are due to the time costs associated with updating the problem state at the end of a step in the respective hard conditions. For instance, after entering a digit in the hard subtraction task, the model updates its problem state to indicate that it finished a step in the subtraction task. The text-entry task only starts when this problem state update is finished, causing a slight delay in the start of the text-entry task.

Alternative strategies

Except for an account based on a problem state bottleneck, there might be other possible explanations for the interaction effects. For example, participants might have employed different task strategies depending on the task condition. However, in the case of the text-entry task it is not easy to come up with alternative strategies because the task is so straightforward. In the easy condition, participants have to read a letter and click a button, which does not seem to allow for multiple strategies. In the hard text-entry condition, participants have to memorize the word, as they do not receive any feedback at all. Furthermore, they have to keep track of where they are within a word, for instance by memorizing the position or the last letter they entered. While the model does memorize the position, alternative strategies exist such as memorizing the last entered letter and reconstruct the position from that information. However, irrespective of which strategy was used, participants will have to keep track of the current position in some way, for which we assume they have to use their problem state.

In the case of the subtraction task there is at least one possible alternative strategy. Participants could have used the display to determine whether or not a borrowing is in progress instead of maintaining a problem state (i.e., looking at the previous subtraction column, if the lower term is higher than the upper term, a borrowing is in progress). If this had been the overall strategy, it would have had the same impact on both the *hard subtraction – easy text-entry* and *hard subtraction – hard text-entry* conditions. In that case, one would not expect to find an interaction effect, as the problem state is not used for the subtraction task. However, it is possible that participants only switched to this strategy in the *hard–hard* condition: thus using a problem state strategy as long as text-entry is easy, and switching to an interface strategy when text-entry became hard. This would incur a time cost in the *hard–hard* condition, and would thus have resulted in a similar interaction effect as we found. To rule out this alternative explanation, we controlled for this in Experiment 3 by masking previous columns, yielding, as we will see, the same results. Obviously, alternative strategies also exist for solving a borrow-in-progress. For instance, one could subtract one from the upper term of the next column, or add one to the lower term, giving the same results. However, in both cases it is necessary to keep track of whether a borrowing is in progress, resulting in similar latency predictions.

Is a problem state bottleneck necessary?

As the threaded cognition theory already proposes a number of bottlenecks, is an additional problem state bottleneck necessary to account for the observed interaction? The over-additive interaction is caused by a resource that is required in both hard conditions, but not in the other conditions. As the hard conditions require additional information to be kept available, a bottleneck should be related to this additional information maintenance. The bottleneck associated with production rule execution cannot offer an explanation for the found interactions, because production rule activity cannot store information without using another resource. A possible alternative explanation is that 'problem states' are stored as declarative memory chunks and are retrieved when needed, instead of having a separate problem state resource. In such a model, however, one would not expect to find an interaction effect because declarative memory is never concurrently required by the two tasks, as the participants have to alternate between the tasks. Thus, in that case the first task would retrieve its problem state from declarative memory and give a response, after which the second task would retrieve its own problem state from declarative memory and give a response, etc. Because declarative memory is in that case never required by both tasks at the same time, it cannot explain the effect of one task on the other task. Thus, we would predict a simple additive effect of conditions, not an interaction effect. As the two peripheral bottlenecks cannot be used to store information, we argue that a problem state bottleneck is the most plausible option to account for the human data.

Cognitive load effects

While we argued above that a problem state bottleneck is the most plausible account within the ACT-R-based threaded cognition theory, there is an extensive psychological literature on cognitive load that can also explain the results of Experiment 1. For instance, it is shown that memory load causes an increase in reaction time in tasks as simple as visual search (e.g., Logan, 1979; Woodman, Vogel, & Luck, 2001) and tone classification (Jolicœur & Dell'Acqua, 1999). Thus, in that sense it is not surprising that maintaining an additional memory load (problem state) influences another task with a memory load, resulting in the over-additive interaction effect. To rule out the possibility of cognitive load causing the interaction effect, Experiment 2 was designed. The dual-task setup of Experiment 1 was slightly modified by requiring the participants to switch tasks only after every two responses in each task. Thus, Experiment 2 also includes responses where no problem state switch is required, but where a memory load representing the state of the other task still has to be maintained (see Figure 2.5). This means that the cognitive load is equal (the memory load of the other task) on both responses, while the problem state only has to be switched for the first response and is still available for the second response. According to a cognitive load account, the interaction effect should be present on both responses, but according to a problem state bottleneck account, the interaction effect should only be present on the first response, and disappear on the second.

Experiment 2: Subtraction & Text-Entry – Two Responses Per Switch

Experiment 2 was designed to test whether the problem state bottleneck can be observed when controlling for cognitive load effects. The design of the experiment was the same as Experiment 1, except that participants now had to give two responses on each task before switching to the other task. Thus, the new experiment has a 2 × 2 × 2 design (Subtraction Difficulty × Text-Entry Difficulty × Switch). Switch responses are the first responses on a task, directly after switching from the other task; non-switch responses are the second responses on a task, following a response in the same task (cf. task switching). Figure 2.5 shows the experimental setup, detailing when a memory / cognitive load is present, and when problem state changes are required in the *hard–hard* condition. On the basis of the problem state bottleneck hypothesis and the outcome of Experiment 1, we predict an over-additive interaction effect in the switch condition (because the problem state has to be replaced for each response in the *hard–hard* condition), but simple additive main effects in the non-switch condition (because the problem state does not have to be replaced in any condition, as the previous response was given in the same task). Because the memory load is the same on switch and non-switch responses (whether a borrowing is in progress for subtraction / what the word and position are for text-entry), a cognitive load account would predict identical effects for both switch and non-switch responses. Thus, we did not introduce additional cognitive load, but merely removed problem state changes on the non-switch responses, enabling the comparison between a cognitive load account and a problem state account.

Method

Participants

Fifteen students of the University of Groningen who did not take part in Experiment 1 participated in the experiment for course credit (9 female, age range 18–23, mean age 19.8). All participants had normal or corrected-to-normal visual acuity. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

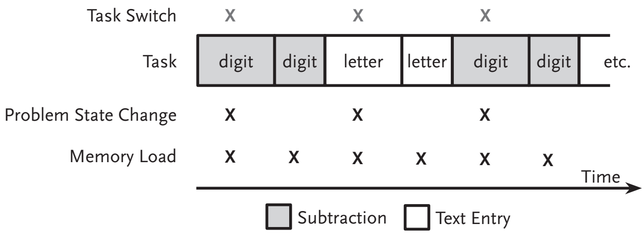


Figure 2.5 Experimental setup of Experiment 2. Grey boxes represent the subtraction task, white boxes the Text Entry task. The black Xs show the problem state and memory load in the hard–hard condition: on the switch responses there is both a problem state change and a memory load, while on the non-switch responses only a memory load is present.

Design, Stimuli, & Procedure

Design, stimuli and procedure were identical to Experiment 1, except that participants were now required to alternate after every two responses, thus they had to enter two digits, two letters, two digits, et cetera.

Model

The model of Experiment 1 was extended to enable it to respond in the situation where a response directly followed a response within the same task.[5] Furthermore, we scaled retrieval times of declarative facts and number of incorrect retrievals to match the new participant group's cognitive arithmetic ability, as we did in Experiment 1.[6]

Results

Only the data of the experimental phase were analyzed. One participant did not adhere to task instructions and was removed from the data set. The same exclusion criteria were used as in Experiment 1 (3.8% of the data was rejected). If not noted otherwise, analyses were the same as in Experiment 1. Figures 2.6 and 2.7 show the main results for response times and accuracy.

Response Times

In line with our hypothesis, ANOVAs on response times showed significant three-way interactions of Subtraction Difficulty × Text-Entry Difficulty × Switch on both the response times of the text-entry task ($F(1,13) = 29.99$, $p = .001$, $\eta_p^2 = .70$) and the subtraction task ($F(1,13) = 5.96$, $p = .03$, $\eta_p^2 = .31$). Therefore, the following analyses were performed separately on the switch and non-switch data.

[5] *While the model was extended, we could have used this new model for Experiment 1 without affecting the results; the situation in which a response can be followed by a response on the same task just never occurs in Experiment 1.*

[6] *ACT-R's latency factor and activation noise were not changed (respectively .3 and .1). The activation levels of the three groups of arithmetic chunks of Footnote 4 were adjusted for the new group of participants.*

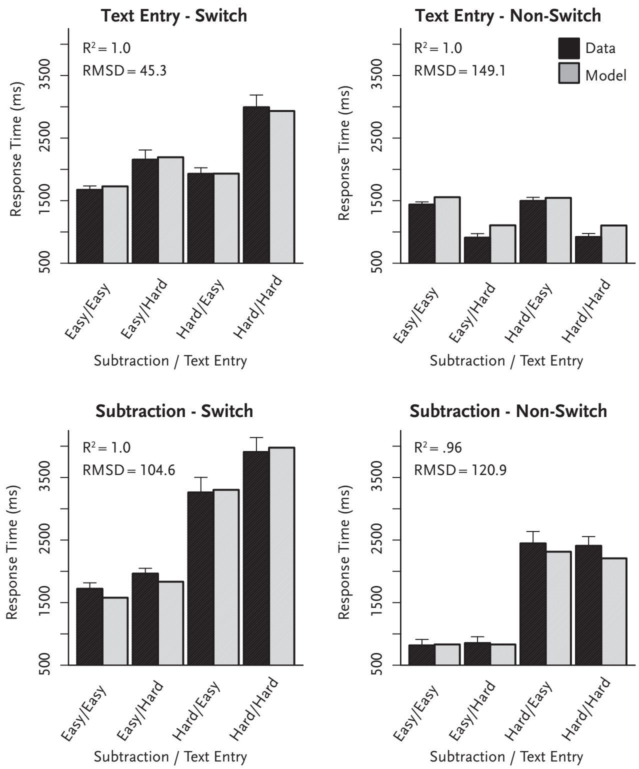


Figure 2.6 Response time data of Experiment 2. RMSD = root mean squared deviation.

The upper panels of Figure 2.6 show the response times on the text-entry task. On the left the switch data are shown, on the right the non-switch data. As predicted, an interaction effect of Subtraction Difficulty and Text-Entry Difficulty was found on the switch trials ($F(1,13) = 19.8$, $p < .001$, $\eta_p^2 = .60$). A simple effects analysis subsequently revealed significant effects of Text-Entry Difficulty when subtraction was easy ($F(1,13) = 27.6$, $p < .01$, $\eta_p^2 = .68$), Text-Entry Difficulty when subtraction was hard ($F(1,13) = 59.2$, $p < .001$, $\eta_p^2 = .82$), Subtraction Difficulty when text-entry was easy ($F(1,13) = 13.2$, $p < .01$, $\eta_p^2 = .50$), and Subtraction Difficulty when text-entry was hard ($F(1,13) = 135.1$, $p < .001$, $\eta_p^2 = .91$). Thus, response times on the switch responses of the text-entry task increased with task difficulty, with an over-additive interaction effect when both tasks were hard. An analysis of the non-switch responses of the text-entry task (upper right panel) showed that only the main effect of Text-Entry Difficulty reached significance, ($F(1,13) = 377.53$, $p < .001$, $\eta_p^2 = .97$). The main effect of Subtraction Difficulty ($F < 1$) and the interaction effect between Subtraction Difficulty and Text-Entry Difficulty ($F(1,13) = 2.4$, $p = .15$, $\eta_p^2 = .16$) were not significant. Note that response times decreased with Text-Entry Difficulty, instead of increasing.

The two lower panels of Figure 2.6 show response times on the subtraction task. The left panel shows the switch responses. An ANOVA revealed a significant interaction effect of Subtraction Difficulty and Text-Entry Difficulty ($F(1,13) = 6.9$,

$p = .02$, $\eta_p^2 = .35$). Subsequent simple effects analyses showed significant effects of Text-Entry Difficulty when subtraction was easy ($F(1,13) = 19.1$, $p < .001$, $\eta_p^2 = .59$), Text-Entry Difficulty when subtraction was hard ($F(1,13) = 14.7$, $p < .01$, $\eta_p^2 = .53$), Subtraction Difficulty when text-entry was easy ($F(1,13) = 104.9$, $p < .001$, $\eta_p^2 = .89$), and Subtraction Difficulty when text-entry was hard ($F(1,13) = 185.5$, $p < .001$, $\eta_p^2 = .93$). Thus, response times on the switch responses of the subtraction task increase with task difficulty, with an over-additive interaction effect, resulting in the highest response times in the *hard–hard* condition. The non-switch response times are shown in the lower right panel of Figure 2.6. Only the main effect of Subtraction Difficulty was significant ($F(1,13) = 305.2$, $p < .001$, $\eta_p^2 = .96$), the main effect of Text-Entry Difficulty and the interaction effect were not significant, Fs < 1. Thus, non-switch response times were lower when the subtraction task was easy.

Accuracy

Figure 2.7 shows the accuracy data of Experiment 2. An ANOVA on the text-entry data shows only a significant main effect of Text-Entry Difficulty ($F(1,13) = 9.7$, $p < .01$, $\eta_p^2 = .43$). The main effect of Switch ($F(1,13) = 2.23$, $p = .16$, $\eta_p^2 = .15$) and Subtraction Difficulty ($F(1,13) = 1.46$, $p = .25$, $\eta_p^2 = .10$) were not significant, neither were the interaction effects between Switch and Subtraction Difficulty ($F < 1$), Switch and Text-Entry Difficulty ($F(1,13) = 3.29$, $p = .09$, $\eta_p^2 = .20$), Subtraction and Text-Entry Difficulty ($F < 1$), and the three-way interaction between Switch, Subtraction Difficulty and Text-Entry Difficulty ($F(1,13) = 3.52$, $p = .08$, $\eta_p^2 = .21$). Thus, accuracy on the text-entry task was lower when text-entry was hard.

Along the same lines, an analysis of the subtraction data only revealed a significant main effect of Subtraction Difficulty ($F(1,13) = 40.7$, $p < .001$, $\eta_p^2 = .76$). The main effects of Switch ($F(1,13) = 2.55$, $p = .13$, $\eta_p^2 = .16$) and Text-Entry Difficulty ($F < 1$) did not reach significance, neither did the interaction effects of Switch and Subtraction Difficulty ($F < 1$), Switch and Text-Entry Difficulty ($F < 1$), Subtraction Difficulty and Text-Entry Difficulty ($F(1,13) = 1.53$, $p = .24$, $\eta_p^2 = .11$), or the three-way interaction between Switch, Subtraction Difficulty, and Text-Entry Difficulty ($F < 1$). Again, subtraction accuracy only decreased when the subtraction task became hard.

Model

The model fits well to the response time data (Figure 2.6, grey bars, R^2- and RMSD-values are shown in the graphs). It shows on the one hand the interaction effects in the switch responses, caused by the problem state replacements for each response, and on the other hand no interaction effects in the non-switch responses. Furthermore, it reflects the decrease in response times on the text-entry task non-switch responses, when text-entry was hard (the reason why the model shows these effects is discussed below). The model also follows the accuracy data closely (Figure 2.7): In general capturing the (non-significant) interaction effects in the *hard–hard* conditions, but slightly over-estimating these effects in the text-entry task.

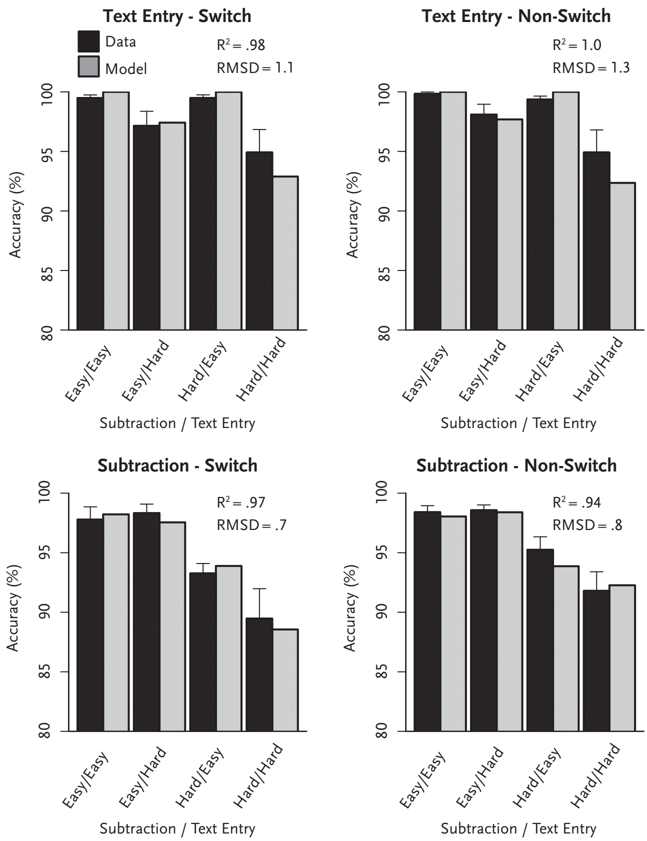


Figure 2.7 Accuracy data of Experiment 2. RMSD = root mean squared deviation.

Discussion

As predicted by the model, an over-additive interaction effect was found on the switch response times in both tasks, but not on the non-switch response times. The model explains this by assuming a problem state bottleneck, requiring the replacement of the problem state in the *hard–hard* condition of the switch responses. In the non-switch responses, on the other hand, the problem state never has to be switched: it is still present from the previous response. A cognitive load account would predict an interaction effect both in the switch and the non-switch responses, because the memory load of the other task is present in both cases (see also Figure 2.5). However, as no interaction effect was observed on the non-switch trials, cognitive load of the other task does not seem to have caused the effects observed in Experiment 1 and in the switch trials in Experiment 2. On the other hand, the model fit shows that a problem state bottleneck account accounts well for the data. Note that we equated cognitive load here with memory load (as for example, Logan, 1979), while there is no consensus in the literature to what exactly constitutes cognitive load. Nonetheless, irrespective of the operationalization of cognitive load, Experiment 2 still gives additional support to a problem state account: When the problem state does not have to be changed, the interaction effect that shows problem state interference disappears.

A second interesting effect is the lower average response time of the hard non-switch text-entry responses, as compared to the easy non-switch responses (upper right panel of Figure 2.6). The model explains this decrease by the fact that in the hard condition it is already known what word has to be entered, and thus also what the next letter is that has to be clicked. Therefore, the model does not have to look at the display of the text-entry task to see what it has to enter, as in the easy version, but can directly search for the correct button and click it. For the switch responses, this decrease in response time is not present, because in that case the model starts the hard text-entry task by retrieving spelling information from declarative memory to determine which letter it has to enter next. On the non-switch responses, the model already initiates the retrieval of the spelling information while clicking the mouse for the previous response, enabling faster responses.

Furthermore, participants were also in general faster on the non-switch responses than on the switch responses. This effect can be explained by the fact that it is necessary to redirect vision and attention to the other task on the other side of the screen on the switch responses, while this is not necessary on the non-switch responses (cf. task switching).

Phonological Loop

Experiment 2 has shown that a memory load probably did not cause the interference effects in the data. However, another possible explanation is that the problem state information in the hard tasks was verbally mediated, and that the phonological loop (e.g., Baddeley & Hitch, 1974) acted as a bottleneck, instead of the problem state resource. That is, if problem state information is rehearsed in the phonological loop, it is possible that there is some overhead in retrieving information when more information has to be rehearsed in the *hard–hard* condition. This alternative account would result in an interaction effect. To test whether the phonological loop is used for storing the problem state information, a third experiment was performed. While Experiment 2 was aimed at the maintenance of the information in working memory without rehearsal, Experiment 3 specifically targets possible rehearsal of the information. In this experiment a listening comprehension task was added to the subtraction and text-entry dual-task, overloading the phonological loop.

Experiment 3: Triple-tasking

For Experiment 3, a listening comprehension task was added to the subtraction and text-entry task: in half of the trials participants had to listen to short stories while performing the other tasks. At the end of a trial, participants had to answer a multiple-choice question about these stories. The experiment has a 2 × 2 × 2 design (Subtraction Difficulty × Text-Entry Difficulty × Listening). Adding a continuous listening task results in the phonological loop being constantly filled with verbal information. If a phonological loop bottleneck was the reason for the interaction effects in Experiments 1 and 2, adding the listening task should produce similar effects in the *easy–hard* and *hard–easy* conditions as we previously saw in the *hard–hard* condition, because now it is also in use by multiple tasks in these conditions. Thus, if the problem state is maintained in the phonological loop, we should now find interference effects as soon as one problem state is stored alongside the information of the listening task (in the *easy–hard* and *hard–easy* conditions). If, on the other hand, the interaction effects were caused by a problem state bottleneck, one would expect the same patterns in the data as found in Experiment 1, with possibly higher response times and error rates over all conditions due to increased cognitive load. Furthermore, our model proposes that, as long as no additional use of the problem state resource is introduced, the problem state bottleneck is independent of the number of tasks and of the amount of cognitive load. Therefore, adding the listening task to the experiment should not influence the results we found previously. To measure baseline performance, the listening task was also tested separately.

Method

Participants

Twenty-three students of the University of Groningen who did not participate in Experiments 1 and 2 participated in Experiment 3 for course credit; one participant had to be excluded because of technical difficulties, resulting in 22 complete datasets (17 female, age range 18–47, mean 22.0). A different set of 6 students participated in the listening baseline experiment (5 female, age range 18–21, mean 19.3). All participants had normal or corrected-to-normal visual acuity and normal hearing. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

Design

The subtraction and text-entry tasks remained unchanged, apart from one thing: columns in the subtraction task that were solved were masked with #-marks, preventing display-based strategies (see the Discussion of Experiment 1). The listening task consisted of listening to a short story during each trial, about which a multiple-choice question was asked at the end of the trial. After answering the question, participants received accuracy feedback, to ensure they kept focusing on the stories. The design of the baseline experiment was similar, but instead of the subtraction and text-entry tasks a fixation cross was shown.

Stimuli

Stimuli for the subtraction and text-entry task were the same as in Experiment 1, except that six additional words were selected. The listening task was compiled out of two official Dutch listening comprehension exams (NIVOR-3.1/3.2, Cito Arnhem 1998). The story length ranged between 17 and 48 seconds (M = 30.4, SD = 10.9). The multiple-choice questions consisted of three options. Two example questions are:

> You would like to buy a new washing machine. When will you get a discount?
> A. If you pay cash.
> B. If you buy an extended warranty.
> C. If you buy a dryer as well.

> You are visiting a laboratory with colleagues. What should you do with your lab coat when you leave?
> A. Put it in the yellow container.
> B. Put it in the green container.
> C. Reuse it.

These questions can be answered without making inferences, but do require attention for the complete duration of the story (i.e., the color of the container in the second question is only said once; participants only see the question after they heard the text).

Procedure

The procedure was identical to Experiment 1 if not noted otherwise. In this experiment, participants had to start each trial with the subtraction task. In the listening condition, playback of the story was initiated simultaneously with the presentation of the subtraction task. Thus, the listening task had to be performed concurrently with the subtraction and text-entry tasks. The multiple-choice question for the listening task was presented either after the feedback screens of the other tasks, or after the story was completely presented, whichever came last. The feedback screen for the listening task was presented for 4 seconds after answering the question. Participants were instructed that the listening task was the most important task, and had to be given priority over the other tasks, while still performing the other tasks as quickly and accurately as possible.

Participants practiced 4 example stories. The experiment consisted of 4 blocks of 12 trials each, 48 trials in total, in a similar setup as Experiment 1. Either the first two blocks were combined with the listening task, or the last two blocks, counterbalanced over participants. The order of the stories was randomized. The complete experiment lasted approximately 60 minutes.

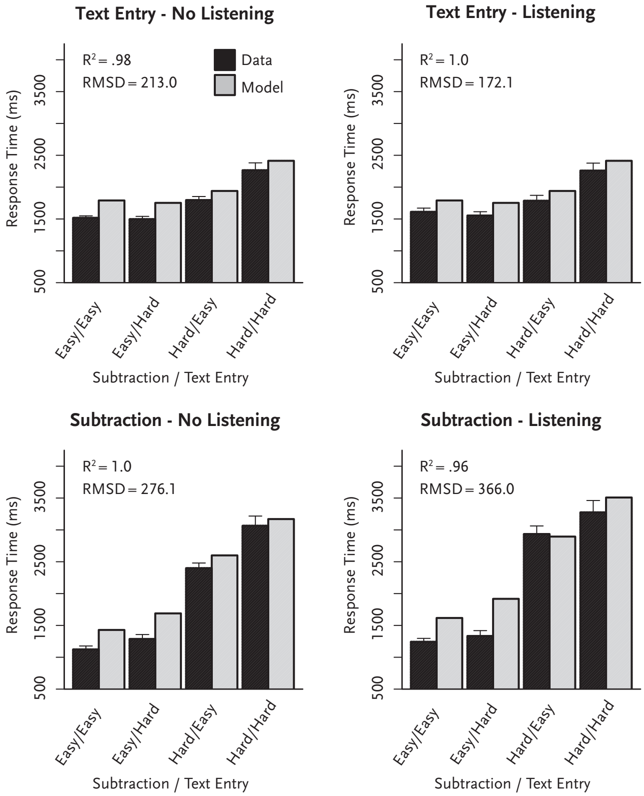


Figure 2.8 Response time data of Experiment 3. RMSD = root mean squared deviation.

Model

The same model as for Experiment 1 was used for the subtraction and text-entry tasks, adjusted for the differences in arithmetic skills between participant groups as was done to calibrate the model to the skill level of the participants in Experiment 2. That is, we adjusted retrieval times of declarative facts and number of incorrect retrievals to match the new group of participants.[7]

To model the listening task, we added a third thread to the model. This thread aurally perceives words, retrieves spelling and syntactic information from memory, and builds simulated syntactic trees. The same approach was used by Salvucci and Taatgen (2008) to model the classical reading and dictation study by Spelke, Hirst, and Neisser (1976), and by Van Rij, Hendriks, Spenader, and Van Rijn (2009) and Hendriks, Van Rijn and Valkenier (2007) to account for developmental patterns in children's ability to process pronouns. This model is a simplified version of Lewis and Vasishth's model of sentence processing (2005; Lewis, Vasishth, & Van Dyke,

[7] *ACT-R's latency factor and activation noise were again left unchanged (respectively .3 and .1). The activation levels of the three groups of arithmetic chunks of Footnote 4 were adjusted for the new group of participants.*

2006) that constructs syntactic trees for sentence processing. For the current model that kind of linguistic detail is unnecessary, as we are mostly interested in how the tasks influence one another. Thus, it suffices to account for the use of procedural and declarative memory in the listening task.

For each word, the aural module processes the word, four procedural rules fire, and two facts are retrieved from memory, which results in about 320 ms processing time per word, which is fast enough to keep up with the speaking rate of 372 ms per word on average.[8] Because ACT-R's aural module is used to perceive the words, using a phonological loop-based strategy is prevented as this strategy is implemented in ACT-R as a combination of the aural and vocal modules (Huss & Byrne, 2003). No control or executive mechanisms were added to the model: the interleaving of the tasks was left to threaded cognition. Answering the multiple-choice questions was not modeled, as this would have required linguistic processing at a level of complexity that is beyond the scope of this paper.

Results

The same exclusion criteria were used as in Experiment 1 (2.4% of the data was rejected). One question from the listening task was removed, as it was consistently answered incorrectly. If not noted otherwise, analyses were the same as in Experiment 1. Because the stories did not always last for the complete trials of the subtraction and text-entry task, some responses on these tasks were made without participants listening to a story. Therefore, we only took responses into account that were made while the story was present.[9]

Response Times

Figure 2.8, upper panel, shows response times on the text-entry task, on the left without and on the right with the listening task. As there is no main effect of Listening, nor any interaction effects involving Listening (all Fs < 1, except for the interaction between Listening and Subtraction Difficulty: $F(1,21) = 1.9$, $p = .18$, $\eta_p^2 = .08$), we collapsed over Listening. The interaction between Text-Entry Difficulty and Subtraction Difficulty was significant ($F(1,21) = 38.78$, $p < .001$, $\eta_p^2 = .65$); a simple effects analysis showed effects of Text-Entry Difficulty when subtraction was hard ($F(1,21) = 37.17$, $p < .001$, $\eta_p^2 = .64$), Subtraction Difficulty when text-entry was easy ($F(1,21) = 30.89$, $p < .001$, $\eta_p^2 = .60$), and Subtraction Difficulty when text-entry was hard ($F(1,21) = 80.60$, $p < .001$, $\eta_p^2 = .79$). Text-Entry Difficulty when subtraction was easy did not reach significance ($F(1,21) = 3.0$, $p = .10$, $\eta_p^2 = .13$). Thus, there was no effect from the listening task on the response times of the text-entry task. Irrespective of the listening task, response times increased

8 *Note that the model is capable of listening to speech faster than 320 ms/word, because the audio module can already start processing the next word while the current word is processed.*

9 *Because this results in an unequal number of observations per cell, we also fitted linear mixed effects models (Baayen et al., 2008). The linear mixed effect models confirmed the ANOVA results. For reasons of consistency, we decided against reporting these additional statistics in the main text, but refer the reader to the Appendix for more details.*

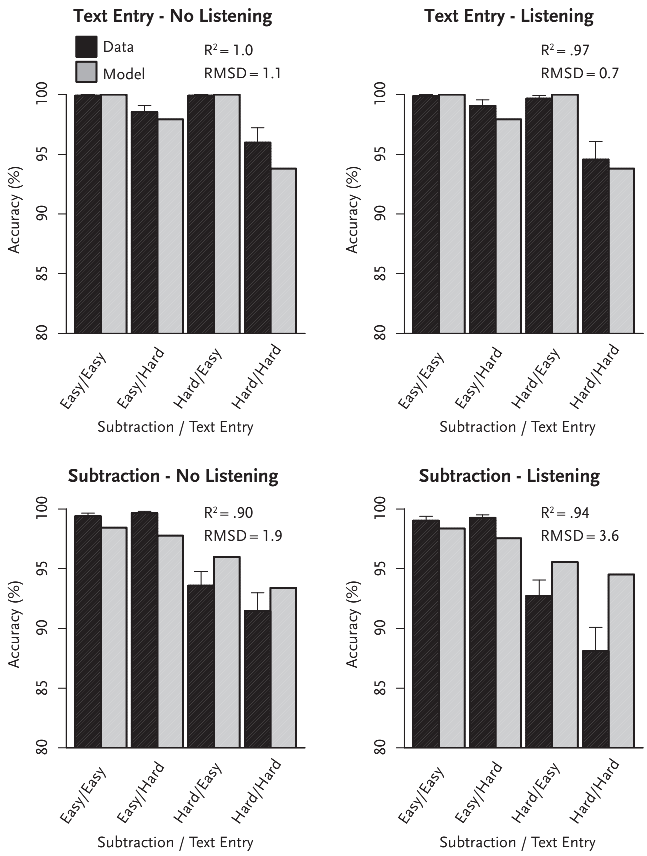


Figure 2.9 Accuracy data of Experiment 3. RMSD = root mean squared deviation.

when subtraction was hard, with an additional increase when text-entry was also hard, resulting in the interaction effect.

The lower panel of Figure 2.8 shows response times on the subtraction task, on the left without, and on the right in combination with the listening task. An ANOVA showed that the three-way interaction between Listening, Subtraction Difficulty, and Text-Entry Difficulty did not reach significance ($F(1,21) = 3.2$, $p = .09$, $\eta_p^2 = .13$), but the main effect of Listening did ($F(1,21) = 4.97$, $p = .04$, $\eta_p^2 = .19$). Furthermore, all two-way interactions reached significance: between Listening and Subtraction Difficulty ($F(1,21) = 9.33$, $p < .01$, $\eta_p^2 = .31$), between Listening and Text-Entry Difficulty ($F(1,21) = 5.98$, $p = .02$, $\eta_p^2 = .22$), and between Subtraction Difficulty and Text-Entry Difficulty ($F(1,21) = 14.3$, $p < .01$, $\eta_p^2 = .40$). A subsequent simple effects analysis of Subtraction Difficulty and Text-Entry Difficulty when the listening task had to be performed revealed significant effects of Text-Entry Difficulty when subtraction was hard ($F(1,21) = 7.12$, $p = .01$, $\eta_p^2 = .25$), Subtraction Difficulty when text-entry was easy

($F(1,21) = 347.1$, $p < .001$, $\eta_p^2 = .94$), and Subtraction Difficulty when text-entry was hard ($F(1,21) = 175.3$, $p < .001$, $\eta_p^2 = .89$). Text-Entry Difficulty when subtraction was easy did not reach significance ($F(1,21) = 3.8$, $p = .07$, $\eta_p^2 = .15$). When the listening task did not have to be performed, all simple effects were significant: Text-Entry Difficulty when subtraction was easy ($F(1,21) = 26.9$, $p < .001$, $\eta_p^2 = .56$), Text-Entry Difficulty when subtraction was hard ($F(1,21) = 27.1$, $p < .001$, $\eta_p^2 = .56$), Subtraction Difficulty when text-entry was easy ($F(1,21) = 337.2$, $p < .001$, $\eta_p^2 = .94$), and Subtraction Difficulty when text-entry was hard ($F(1,21) = 226.7$, $p < .001$, $\eta_p^2 = .92$). Furthermore, Listening had a significant effect when both subtraction and text-entry were easy ($F(1,21) = 4.37$, $p = .05$, $\eta_p^2 = .17$) and when subtraction was hard and text-entry easy ($F(1,21) = 10.5$, $p < .01$, $\eta_p^2 = .33$), but not when subtraction was easy and text-entry was hard or when both tasks were hard (Fs < 1). To summarize, response times on the subtraction task increased when the listening task had to be performed and with task difficulty of the subtraction and text-entry tasks. Furthermore, the effects of Text-Entry Difficulty were smaller when the listening task had to be performed, while the effects of Subtraction Difficulty were larger when the listening task had to be performed (as shown by the two-way interaction effects). An over-additive interaction effect of Subtraction Difficulty and Text-Entry Difficulty was present, both when the listening task had to be performed and when it did not have to be performed.

Accuracy

In Figure 2.9 the accuracy data of Experiment 3 is displayed. The upper panels show the accuracy on the Text-Entry task. As there was neither an effect of Listening, nor any interaction effects involving Listening (all Fs < 1), we collapsed over Listening. The subsequent ANOVA showed an interaction effect of Subtraction Difficulty and Text-Entry Difficulty ($F(1,21) = 6.55$, $p = .02$, $\eta_p^2 = .24$). Three of the four simple effects were significant: Text-Entry Difficulty when subtraction was easy ($F(1,21) = 7.81$, $p = .01$, $\eta_p^2 = .27$), Text-Entry Difficulty when subtraction was hard ($F(1,21) = 33.1$, $p < .001$, $\eta_p^2 = .61$), and Subtraction Difficulty when text-entry was hard ($F(1,21) = 16.0$, $p < .001$, $\eta_p^2 = .43$). Subtraction Difficulty when text-entry was easy did not reach significance ($F < 1$). Thus, accuracy on the text-entry task was lower when text-entry was hard, with an over-additive effect when subtraction was hard as well.

The lower panels of Figure 2.9 show the accuracy data on the subtraction task. Again, there were no significant effects involving Listening (all Fs < 1, except for the main effect of Listening: $F(1,21) = 1.91$, $p = .18$, $\eta_p^2 = .08$), thus we collapsed over Listening. The ANOVA showed a significant interaction effect of Subtraction Difficulty and Text-Entry Difficulty ($F(1,21) = 6.6$, $p = .02$, $\eta_p^2 = .24$). Three simple effects reached significance: Subtraction Difficulty when text-entry was easy ($F(1,21) = 47.2$, $p < .001$, $\eta_p^2 = .69$), Subtraction Difficulty when text-entry was hard ($F(1,21) = 127.4$, $p < .001$, $\eta_p^2 = .86$), and Text-Entry Difficulty when subtraction was hard ($F(1,21) = 10.9$, $p < .01$, $\eta_p^2 = .34$). Text-Entry Difficulty when subtraction was easy was not significant ($F < 1$). Thus, accuracy on the subtraction task was lower when subtraction was hard, and even lower when text-entry was hard as well.

Figure 2.10, left panel, shows the accuracy data of the listening task. The leftmost bar shows the results of the listening baseline experiment (i.e., participants only performed the listening task): 89% correct. Adding the other tasks had little effect, except when both the subtraction and the text-entry task were hard. The interaction between Subtraction Difficulty and Text-Entry Difficulty was significant ($F(1,21) = 7.42$, $p = .01$, $\eta_p^2 = .26$); as were the simple effects of Text-Entry Difficulty when subtraction was hard ($F(1,21) = 9.18$, $p < .01$, $\eta_p^2 = .30$) and Subtraction Difficulty when text-entry was hard ($F(1,21) = 14.75$, $p < .001$, $\eta_p^2 = .41$), driving the interaction effect. The simple effects of Text-Entry Difficulty when subtraction was easy ($F(1,21) = 1.73$, $p = .20$, $\eta_p^2 = .08$) and Subtraction Difficulty when text-entry was easy ($F < 1$) were not significant.

Model

As can be seen in Figure 2.8, the response times of the cognitive model fit well to the human data, both in combination with and without the listening task (R^2 and RMSD values are shown in the graphs). The accuracy data in Figure 2.9 is also accounted for, especially in the text-entry task the model follows the data closely. For the subtraction task the effects are slightly under-predicted: the effects in the data are larger, especially when the listening task is present.

The right panel of Figure 2.10 shows the percentage of words processed by the model. The model can only process words when declarative memory is available. Thus, when words are presented while declarative memory is in use by the other tasks, words cannot be processed, and will be substituted by new words entering the auditory buffer. This happens most often in the *hard–hard* condition, as problem states have to be retrieved from declarative memory for the other tasks on each step of a trial, blocking the resource. Obviously, a percentage of processed words cannot be translated directly into number of correctly answered questions, but the model shows a similar pattern of performance ($R^2 = .68$).

Discussion

In Experiment 3 we added a listening comprehension task to the two tasks used in the previous experiments. The same interaction effects were found as in Experiments 1 and 2, both when the listening task was present and when it was not. Experiment 3 was designed to test whether a problem state bottleneck caused the interference effects, as opposed to a phonological loop bottleneck. If it was a phonological loop bottleneck that caused the interference, overloading the phonological loop by adding the listening task should cause interference effects not only in the *hard–hard* condition, but also in the *hard–easy* and *easy–hard* conditions of the other tasks. The only effect we found that pointed in this direction was the increase of reaction times of the subtraction task when subtraction was hard and the listening task had to be performed (Figure 2.8, lower panel). However, this effect is accounted for in the model by declarative memory interference instead of phonological loop or problem state interference (see below). As the other three conditions did not increase in reaction times, this implies that the

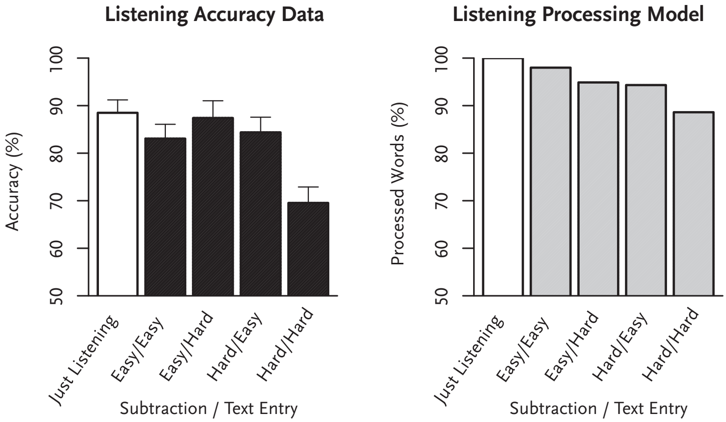


Figure 2.10 Accuracy on the listening task of Experiment 3, and the percentage of processed words by the model per condition.

phonological loop did not cause the interference in Experiment 1 and 2, and provides additional support to a problem state account.

Generally speaking, the listening task had surprisingly little influence on the subtraction and text-entry tasks: the response times only increase by a small amount in the subtraction task. Interestingly, while we did not think about this effect beforehand, and did not model it explicitly afterwards, this increase in response times emerged naturally from our model. A close inspection of the model revealed that it is caused by the continuous use of declarative memory by the listening task. Threaded cognition causes the tasks to be closely interleaved, which means that most of the time there is little interference. However, when the subtraction task needs to use declarative memory when it is in use by the listening task, this will cause a slight delay in execution, causing the increase in response times. This effect is more pronounced when the subtraction task was hard, as shown by the two-way interaction between Listening and Subtraction Difficulty. The model explains this by the need for more declarative retrievals in the hard subtraction task as compared to the easy subtraction task, which leads to more interference with the declarative retrievals of the listening task. For the text-entry task a similar effect would be expected, except for the fact that the text-entry task is much less memory intensive (i.e., less memory retrievals have to be performed) than the subtraction task. That is why the model does not predict an increase in reaction times for the text-entry task, which is consistent with the human data.

The listening task was involved in one more effect: the interaction effect between Listening and Text-Entry Difficulty on the response times of the subtraction task (Figure 2.8, lower panel). That is, the effect of Text-Entry Difficulty was smaller when the listening task had to be performed. An opposite effect would have been expected when the interference effects of Experiment 1 and 2 were caused by a phonological loop bottleneck, because in that case, the phonological loop would have caused interference in combination with only one hard task, as explained above. The current model does not account for the interaction between Listening and Text-Entry Difficulty that was observed in the data. However, as this paper focuses on the effects of the problem state manipulations, and the purpose of the experiment was to rule out any auditory loop related explanation (which would have caused an opposite effect), we decided against adding a post-hoc explanation to the model.

The effects of the subtraction and text-entry task on listening comprehension were also surprisingly small: a decrease in listening accuracy scores was observed only when both other tasks were hard. The model explains this finding by the assumption that declarative memory is in high demand by the subtraction and text-entry tasks when both these tasks are hard, because problem states have to be retrieved from declarative memory on each step of a trial. Therefore, a word is sometimes replaced by the next presented word in the auditory buffer before it is processed using declarative memory. As there is not sufficient time in the *hard–hard* condition to process all words, this will presumably result in more mistakes on the listening comprehension task in this condition.

In conclusion, our threaded cognition model proposed that adding this particular third task should not influence the results of the other tasks dramatically. This turned out to be the case, even while the continuous listening task is, arguably, quite demanding. The patterns in the data were comparable to the data of the previous experiments, while the small increase in response times was explained by the increased use of declarative memory.

General Discussion

In this paper, we tested the hypothesis that the problem state resource acts as a bottleneck in multitasking. Experiment 1 consisted of two tasks that had to be carried out concurrently, both with and without a problem state. This resulted in an over-additive interaction effect of task difficulty (i.e., the requirement of two problem states led to higher response times), confirming the hypothesis. In Experiments 2 and 3, we tested whether this interaction effect was due to cognitive load or to a phonological loop bottleneck, respectively, instead of to a problem state bottleneck. Experiment 2 showed that the interaction effect is not due to a simple memory load effect, but instead is related to a switch of task context. This corroborates the problem state hypothesis. In Experiment 3, the phonological loop was overloaded by adding a story comprehension task. This did not have a major influence on the effects found in Experiment 1, lending additional support to a problem state bottleneck account of the data. Based on these three experiments and general ACT-R assumptions about memory, modularity and performance, we conclude that the problem state resource indeed acts as a bottleneck when it has to be used by multiple tasks concurrently.

Nevertheless, it should be possible to formulate alternative models explaining these data sets, and we therefore cannot claim the data prove the existence of a problem state bottleneck. The strength of the current account over any post-hoc fit of the data is that we tested an a priori prediction made by the threaded cognition theory before running the experiment. First, we ran Experiment 1 to test a qualitative prediction of a problem state bottleneck. Without having to add additional assumptions to the model, it accounted for the interference effects that we found. Subsequently, we tested the two most plausible alternative accounts of the data in Experiments 2 and 3. With the same basic model as used for Experiment 1, we were able to account for the data of these experiments. Thus, based on a theory-driven model we were able to predict the effects of Experiment 1, and could subsequently account for data of two related experiments.

The Problem State and Working Memory

There is a relation between the problem state and the classical notion of working memory (e.g., Baddeley, 1986; Baddeley & Hitch, 1974): both are used for temporarily maintaining mental representations. In the ACT-R architecture, working memory does not exist as a separate system. Instead, working memory is represented by a combination of (a) the contents of the declarative memory buffer and the problem state buffer and (b) highly active chunks in declarative memory (Anderson, 2005; Daily, Lovett, & Reder, 2001; Lewis & Vasishth, 2005; Lovett, Daily, & Reder, 2000). In this scheme, the buffer contents are accessible at no time cost, and thus constitute directly accessible 'true working memory': the 'focus of attention' (e.g., Cowan, 1995; Garavan, 1998; Oberauer, 2002). With a size of two this is comparable to theories positing an extremely limited working memory size (e.g., Garavan, 1998; McElree, 2001). On the other hand, highly active chunks in declarative memory are accessible, but at a small time cost. If these items have just been added, as in common working memory or immediate memory experiments, the number that can be reliably retrieved is around 4 to 9 (e.g., Anderson et al., 1998; Oberauer & Lewandowsky, 2008). This is more comparable to theories with a working memory size of 4 to 9 (e.g., Cowan, 2000; Miller, 1956; Morey & Cowan, 2004). The combination of (a) having a small amount of directly accessible items and (b) a number of easily accessible items at a small time cost, is in line with a number of recent theories (e.g., Jonides et al., 2008; McElree, 2001; Oberauer, 2002).

The problem state acts in this framework as the location where new information is stored. This new information can either originate from perceptual processes (the to-be-entered word in the text-entry task), from a result of processing existing information (the carry-flag in the subtraction task), or from changing existing information (for example, processing $2x + 5 = 8$ to $2x = 3$). This is potentially important for many dual- or multitask situations, as it is often necessary to maintain new information. As we have argued that the problem state acts as a bottleneck, this could have a considerable influence on tasks in which multiple sets of information have to be maintained at the same time.

Threaded Cognition

While our hypothesis was inspired by the threaded cognition theory, one could wonder whether threaded cognition is a necessary part of the models. It is indeed possible to think of a way of modeling our experiments using only ACT-R. The merit of threaded cognition, though, is that we were not forced to come up with a supervisory control structure to model the tasks.[10] This would have been possible, as is best shown by existing multitasking models without threaded cognition (e.g., Anderson, Taatgen, & Byrne, 2005; Salvucci, 2006; Taatgen, 2005). However, these models represent all tasks in a single goal representation. This is hard to defend if the tasks in the experiment are tasks that the participants are already proficient in, like driving or multi-column subtraction. The importance of using threaded cognition is that the existing threads can be reused if, for instance, the subtraction task would be combined with driving. This seems to be the way humans would handle this: our participants would seemingly not have to learn a new supervisory control structure if they had to solve subtraction problems while driving (see, for a more in-depth discussion of this issue, Kieras et al., 2000; Salvucci, 2005; Salvucci & Taatgen, 2008). Case in point is our Experiment 3, in which we were able to add an additional thread for the listening task, without having to change anything in the existing model. The interaction of the three threads, without any supervisory control, turned out to be a good predictor of the participants' behavior: Even the slight increase in reaction times in the listening condition was accounted for by the model, while not predicted by the authors beforehand.

One potential criticism of threaded cognition is that it allows for an unlimited set of goals that is not susceptible to decay. Altmann and Trafton (2002) have successfully argued against the construct of the goal stack, which had the same problem. Instead, Altmann et al. (Altmann & Gray, 2008; Altmann & Trafton, 2002) have proposed that only a single goal can be active at a time, and multiple goals have to be handled by swapping out the current goal with goals retained in declarative memory. Salvucci and Taatgen (2008), however, found that such a procedure would be too slow to account for certain Psychological Refractory Period experiments. The model we have presented here is consistent with both the Salvucci and Taatgen approach, as well as with the Altmann et al. approach. Instead of swapping out the goal as such, though, the contents of the problem state resource are swapped out. The main difference is our assumption that not all tasks require a problem state. Although we have not applied the strategic encoding strategies that Altmann et al. use in their models, this would certainly be possible if a task would necessitate it (see also Salvucci, Taatgen, et al., 2009).

Single vs. Multiple Bottlenecks

In this article we have introduced the notion of multiple bottlenecks, and it is therefore useful to contrast it with a single bottleneck approach (e.g., Pashler, 1994). Explaining multitasking interference with multiple bottlenecks can be considered as a refinement of a single bottleneck account. Single-bottleneck accounts consider central cognition as a uniform system that can only be engaged in a single action at a time. Although this offers accurate accounts of many combinations of simple tasks, the more complex tasks discussed in this article need a more refined theory. Multiple-bottleneck models

[10] *One could argue that executive control plays a role in threaded cognition, the threads act in a 'greedy and polite' way after all. However, this is a task-unspecific form of executive control, not customized for the tasks at hand, and not influencing the interleaving of the tasks directly.*

allow parallel processing of certain combinations of tasks, as long as they use different central resources. In Experiment 3, for example, the listening task almost continuously engages central cognition, but because central cognition is subdivided into separate resources it is still possible to do the other tasks by properly interleaving them, resulting in only a minor impact on performance.

The three bottlenecks or resources we focus on in this article have different impacts on performance, which is mainly due to the time scale on which they operate. Interference in the fastest system, the procedural resource is usually very limited and in the order of tens of milliseconds, and therefore hardly noticeable in the experiments discussed here (although it is in perfect time sharing experiments, Salvucci & Taatgen, 2008, and attentional blink experiments, Taatgen et al., 2009). Interference in the declarative memory resource is usually limited to the maximum duration of a memory retrieval, which is never more than a couple of hundreds of milliseconds in our experiments. This produces small amounts of interference, especially noticeable in the listening task in Experiment 3. The problem state resource, finally, can produce considerable interference, because threads need this resource over longer periods of time. Using threaded cognition it is possible to predict quantitatively how much two tasks will interfere with each other. Single bottleneck models usually do not deal with experiments in which a problem state needs to be maintained over longer periods of time, but nevertheless the problem state behaves like a bottleneck in the same way as procedural bottlenecks in for example perfect time-sharing experiments. We therefore do not see multiple bottlenecks as a refutation of the single bottleneck theory, but rather as a refinement in the details and an extension in time scale.

Implications of a Problem State Bottleneck

Why is the problem state bottleneck important for real life situations? A clear example can be found in our previous research, in which participants had to steer a simulated car and operate a navigation device at the same time (Borst & Taatgen, 2007). It was shown that as soon as participants had to use a problem state for both tasks, their performance decreased considerably. This can be tied back to real life: as soon as information is not readily available in the world, performance levels will decrease if two tasks require the maintenance of intermediate information. Thus, it is preferable to have at most one task that requires the use of a problem state in a multitasking situation. As a design guideline this means that, for example in cars, a secondary device should present its information to the user, instead of requiring the user to maintain intermediate representations.

However, if there is an ongoing task that requires the use of problem representations, and it is known that it will be interrupted (including self-interruptions), human-computer interface designers should try to ensure that the task is interrupted at a point without a problem state. If that is not possible, the user should at least be given the opportunity to rehearse the problem state before the task is suspended. For example, when your work is interrupted by a phone call, most people would let the telephone ring a couple of times before picking it up, and only interrupt their work at a point where it is easy to resume it afterwards. Trafton, Altmann, Brock, and Mintz (2003) formally showed this effect: if users were warned 8 seconds before their task was interrupted, they were significantly faster in resuming the original task than users who were interrupted without a warning. According to our problem state bottleneck theory, the warning gave users the opportunity to rehearse their problem state before being interrupted, while that was impossible in the non-warning condition, enabling faster resumptions after the interruptions (see Salvucci, Taatgen, et al., 2009, for simulations of this experiment).

Conclusion

In summary, the three experiments showed that the problem state resource acts as bottleneck in multitasking. Because the intermediate representations that are stored as a problem state often have to be maintained for several seconds or more, this bottleneck can result in considerable interference between tasks, and therefore has to be taken into account when designing environments for multitasking.

Appendix: More Detailed Analysis of Experiment 3

In this appendix we discuss an alternative analysis of the data of Experiment 3. Because the stories in Experiment 3 did not always last for the complete trials of the subtraction and text-entry task, some responses on these tasks were made without participants listening to a story. Therefore, we only took responses into account that were made while the story was present. Because this results in an unequal number of observations per cell, we also fitted linear mixed effects models to analyze the data (e.g., Baayen, Davidson, & Bates, 2008). As the results of the linear mixed effect models are very similar to those of the ANOVAs, we have included the ANOVA results in the body of the manuscript for reasons of consistency. The results of the linear mixed effects models are reported here.

Response Times

Figure 2.8, upper panels, shows response times on the text-entry task, on the left without and on the right in combination with the listening task. A linear mixed effects model was fitted to the response time data, with Listening, Subtraction Difficulty, and Text-Entry Difficulty as fixed effects and Subject as a random effect. The model shows significant contributions of Listening ($\beta = 88.83$, $t(7836) = 2.84$, $p < 0.01$), Subtraction Difficulty ($\beta = 278.1$, $t(7836) = 9.30$, $p < 0.001$), the Listening × Subtraction Difficulty interaction ($\beta = -142.7$, $t(7836) = -3.03$, $p < 0.01$), and of the Subtraction Difficulty × Text-Entry Difficulty interaction ($\beta = 489.7$, $t(7836) = 11.57$, $p < 0.001$). Comparing this model to a model without the interaction between Subtraction Difficulty and Text-Entry Difficulty shows that the first model is to be preferred ($\chi^2(2) = 206.1$, $p < 0.001$), indicating a significant contribution of the interaction term. Thus, response times were higher when the listening task had to be performed and when the subtraction task was hard, while the combination of the listening and the hard subtraction task caused response times to decrease. Most importantly, an over-additive interaction effect of Subtraction Difficulty and Text-Entry Difficulty was found, irrespective of the listening task. The difference with the ANOVA results reported in the main text is the small influence of the listening task on response times. Although this would argue against collapsing over the Listening conditions for the ANOVA, collapsing over Listening does not change the main outcome of the analysis, nor does it influence our conclusions. Therefore, we opted to keep the main text a consistent whole, and collapsed over Listening.

The lower panels of Figure 2.8 show response times on the subtraction task. Again, a linear mixed effects model was fitted to the data. Listening, Subtraction Difficulty, and Text-Entry Difficulty were added as fixed effects, while Subject was entered as random effect. All main effects contributed significantly to the response times: Listening ($\beta = 112.5$, $t(7854) = 2.51$, $p = 0.01$), Subtraction Difficulty ($\beta = 1276$, $t(7854) = 29.37$, $p < 0.001$), and Text-Entry Difficulty ($\beta = 164.8$, $t(7854) = 3.82$, $p < 0.001$), as did all interaction effects except Listening × Text-Entry Difficulty: Listening × Subtraction Difficulty ($\beta = 344.0$, $t(7854) = 5.1$, $p < 0.001$), Subtraction Difficulty ×

Text-Entry Difficulty ($\beta = 494.8$, $t(7854) = 8.0$, $p < 0.001$), and the three-way interaction of Listening × Subtraction Difficulty × Text-Entry Difficulty ($\beta = -356.9$, $t(7854) = -3.68$, $p < 0.001$). A comparison between this model and a model without the Subtraction Difficulty × Text-Entry Difficulty interaction showed that the first model fits better to the data ($\chi^2(2) = 67.34$, $p < 0.001$), indicating a significant contribution of the Subtraction Difficulty × Text-Entry Difficulty interaction. Comparing this model to a model without the three-way interaction showed that the first model again fits the data better ($\chi^2(1) = 13.1$, $p < 0.001$), thus, also the three-way interaction contributes significantly to the model. This means that response times were higher when the listening task had to be performed, when the subtraction task was hard, and when the text-entry task was hard; and that the effect of subtraction difficulty is larger in the presence of the listening task. Furthermore, there is a significant interaction between Subtraction Difficulty and Text-Entry Difficulty, which is larger without the listening task than with the listening task. The main difference with the results of the ANOVA is the significant three-way interaction between Listening, Subtraction Difficulty, and Text-Entry Difficulty. The three-way interaction shows that the effect of the interaction between Subtraction Difficulty and Text-Entry Difficulty is smaller when the listening task has to be performed. However, even in the presence of the three-way interaction, the two tasks still interact, which is in accordance with our modeling results.

Accuracy

In Figure 2.9 the accuracy data of Experiment 3 is displayed. The upper panel shows the accuracy on the text-entry task. A binomial linear mixed effects model was fitted to the data with Listening, Subtraction Difficulty, and Text-Entry Difficulty as fixed effects, and Subject as a random effect. It shows only a significant effect of Text-Entry Difficulty ($\beta = -2.9$, $z(7836) = -2.74$, $p < 0.01$). Thus, accuracy on the text-entry task was lower when the text-entry task was difficult. The ANOVA reported in the main text also found a significant interaction between Subtraction Difficulty and Text-Entry Difficulty.

The lower panels of Figure 2.9 show the accuracy data on the subtraction task. A binomial mixed effects model with Listening, Subtraction Difficulty, and Text-Entry Difficulty as fixed effects and Subject as a random effect showed that only Subtraction Difficulty contributes significantly to the model ($\beta = -2.4$, $z(7854) = -6.1$, $p < 0.001$). Thus, accuracy on the subtraction task decreased with Subtraction Difficulty. Again, the ANOVA also found a significant interaction effect of Subtraction Difficulty and Text-Entry Difficulty.

Figure 2.10, left panel, shows the accuracy data of the listening task. The leftmost bar shows the results of the listening baseline experiment (i.e., participants only performed the listening task): 89% correct. Adding the other tasks had little effect, except when both the subtraction and the text-entry task were hard. Fitting a binomial linear mixed effects model with Subtraction Difficulty and Text-Entry Difficulty as fixed effects and subject as random effect, shows a significant interaction effect between Subtraction Difficulty and Text-Entry Difficulty ($\beta = -1.2$, $z(507) = -2.55$, $p = 0.01$). However, because the stories lasted sometimes longer than the other two tasks, parts of the stories were attended without performing the other tasks. If we add the proportion overlap between the stories and the other tasks to the linear model, this does not significantly improve the first model ($\chi^2(4) = 3.74$, $p = .44$). Thus, adding the overlap did not change the outcome of the analysis, leaving only a significant decrease in accuracy when both the subtraction and the text-entry task were hard.

Bypassing the Problem State Bottleneck

In which we show that the problem state bottleneck causes an increase in mental workload and that it can be bypassed by presenting information in the environment.

The experimental data in this chapter was previously published in:

Buwalda, T.A., Borst, J.P., Taatgen, N.A., & Van Rijn, H. (2011). Evading a multitasking bottleneck: Presenting intermediate representations in the environment. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Bypassing the Problem State Bottleneck

Chapter 3

Abstract

Objective: *In this paper we investigate whether external support can prevent negative effects of the problem state bottleneck in human multitasking.*

Background: *Previously, it was shown that the problem state resource – a central element in working memory that maintains current task information – can only be used for one task at a time. When the problem state resource is required for multiple tasks concurrently, performance decreases.*

Method: *To see whether external support reduces the effects of the problem state bottleneck, we measured performance and pupil dilation (to assess mental workload) during an experiment that manipulated the use of the problem state resource.*

Results: *It was shown that the effects of the problem state bottleneck on response times and accuracy diminished when problem state information was presented externally. However, we did not find a difference in mental effort between the two conditions. A cognitive model was used to show that the participants behaved rationally: they only used the information in the environment when more than one problem state was required to do the task, thus only when it improved their performance.*

Conclusion: *We conclude that external support can be used to bypass the problem state bottleneck, but that external support is only beneficial when multiple problem states are required to do a task.*

Application: *These results should be taken into consideration when designing interfaces and tasks: users should at most need a single mental representation to carry out a task. Otherwise, response times and number of errors will increase.*

Introduction

Multitasking is all around us. González and Mark (2004) have shown that people switch on average every 3 minutes between tasks in a typical office environment. In addition, a recent study showed that every generation 'multitasks' more than the previous generation in their free time (Carrier et al., 2009). However, it is also well known that performance on individual tasks suffers from multitasking. In the field of sequential multitasking (i.e., switching between tasks, Salvucci, Taatgen, et al., 2009), theorists have focused on the disruptive effects of interruptions (e.g., Gillie & Broadbent, 1989; Monk et al., 2008). Likewise, the concurrent multitasking literature has identified several processing bottlenecks that lead to decreased performance when two tasks are performed at the same time (e.g., Broadbent, 1958; Keele, 1973; Pashler, 1994; Salvucci & Taatgen, 2008; Wickens, 1984, 2002). One cause of multitasking interference, both in concurrent and sequential multitasking, is the problem state bottleneck (Borst, Taatgen, & Van Rijn, 2010).

The problem state is defined as the element in working memory that can be used without any time cost (Anderson, 2005), unlike other elements in working memory (see e.g., McElree, 2001). It is used to represent intermediate information in a task, for example, '$3x = 15$' when solving '$3x - 5 = 10$'. Previously, we have shown that the problem state resource can contain at most one chunk of information, and therefore causes multitasking interference when required by multiple tasks at the same time (Borst, Taatgen, & Van Rijn, 2010). In a dual-task paradigm, participants needed a problem state for none, one, or both of the tasks. In the condition where subjects needed a problem state for both tasks, performance decreased considerably both in reaction times and accuracy as compared to the other conditions. Supported by a cognitive model, this was taken as an indication of a problem state bottleneck. Further evidence for a problem state bottleneck was provided by Salvucci and Bogunovich (2010), who showed that when subjects had to switch between an e-mail and a chat task, they chose switch points at which they did not have to maintain a problem state.

Given that the problem state bottleneck can lead to a decrease in performance, both in laboratory and real-life settings, we investigated how this bottleneck can be bypassed. In this article we describe an experiment in which participants had to perform two tasks at the same time. The first condition of the experiment is a replication of our previous study (Borst, Taatgen, & Van Rijn, 2010), and should result in problem state interference. In the other condition, we presented supporting information on the screen, thereby offloading possible internal representations to the environment (e.g., Kirsh, 1995). We hypothesized that the interference effects disappear in this condition. A second question that we address in this article is whether the problem state bottleneck causes an increase in mental workload, and whether this possible increase disappears with external support. To assess the level of mental workload during the experiment we measured pupil dilation (e.g., Beatty, 1982). Finally, to show that a problem state bottleneck can account for the observed behavior, we present a computational cognitive model. In the remainder of this article, we first describe the used methodology, followed by the results of the experiment, the model description and results, and a general discussion.

Method

In the experiment, based on earlier experiments by Borst, Taatgen, and Van Rijn (2010), participants had to perform two tasks concurrently: a subtraction task and a text-entry task. Both tasks were presented in two versions: an easy version in which there was no need to maintain a problem state, and a hard version in which participants had to maintain a problem state from one response to the next. In the current paper we extended the original setup with a condition in which the problem state of the subtraction task is displayed on the screen (the support condition), reducing the need for mentally maintaining a problem state in the hard subtraction condition. Thus, the experiment had a 2 × 2 × 2 factorial within-subjects design (Subtraction Difficulty × Text-Entry Difficulty × Support).

Pupil Dilation

To assess mental workload, pupil dilation was measured throughout the experiment. Since the 1960s, pupil size is known to reflect mental activity (e.g., Hess & Polt, 1964) and memory load (e.g, Kahneman & Beatty, 1966). In 1982, based on a large body of research, Beatty argued that pupil dilation could be used as a physiological measure of mental effort, because it reflects "within-task, between task, and between-individual variations in processing demands" (Beatty, 1982, p. 276; see for a more recent review, Steinhauer & Hakerem, 1992). From an applied perspective, Iqbal and colleagues used pupil dilation to study mental workload in a route planning and in a document editing task (Iqbal, Adamczyk, Zheng, & Bailey, 2005). The use of pupil dilation allowed them to track mental workload throughout the tasks, and identify opportunities to interrupt users on points of low workload. In the current task we measured pupil dilation to see if the decrease in performance when participants have to use multiple problem states concurrently is linked to an increase in mental effort, and if this possible increase disappeared when participants receive external support in the subtraction task.

Participants

Thirty-three students of the University of Groningen participated in the experiment for course credit or monetary compensation of €10. Four participants were rejected because they scored less than 75% correct where the other participants scored >95% correct. Two participants were rejected because they did not adhere to task instructions, and three because of recording problems of the eye tracker. This leaves 24 complete datasets (17 female, age range 18-43, mean age 20.5). All participants had normal or corrected-to-normal visual acuity. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

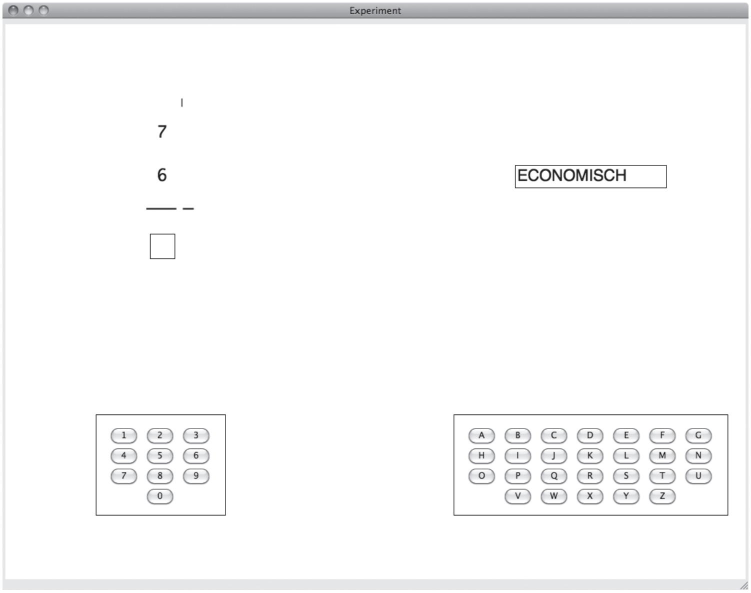


Figure 3.1 The experiment in the support condition. The '|' indicates that there is currently no carry, it will turn into a '1' when a carry has to be processed. Note that in the real experiment one of the tasks would be disabled at any given moment.

Design

During the experiment, participants had to perform a subtraction task and a text-entry task concurrently. The subtraction task was shown on the left side of the screen, the text-entry task on the right (see Figure 3.1). Participants had to alternate between the two tasks: after entering a digit, the subtraction interface was disabled, forcing the participant to subsequently enter a letter. After entering a letter, the text-entry interface was disabled and the subtraction interface became available again.

The subtraction task is shown on the left side of Figure 3.1. Participants had to solve 10-column subtraction problems in standard right to left order. However, at each point in time, only a single column was visible. Although the problems were presented column by column, the participants were instructed that the separate columns in a trial were part of a 10-column subtraction problem (in the practice phase participants started out with a normal 10-column layout, only later they switched to solving the problems column by column). Participants had to enter the digits by clicking on the on-screen keypad with the mouse. In the easy, no problem state version, carrying was never needed because the upper digit was always larger or equal to the lower one. In contrast, the hard version required participants to carry six times out of 10 possible columns. The assumption is that participants use their problem state resource to store whether a carry is in progress.

The interface for the text-entry task is shown on the right in Figure 3.1. Participants had to enter 10-letter strings by clicking on the on-screen keyboard. In the easy version these strings were presented one letter at a time and participants had to click the corresponding button on the keyboard. In the hard version, a 10-letter word was presented once at the start of a trial. Once a participant clicked on the first letter, the word disappeared and the remaining letters had to be entered one at a time, without feedback. Thus, after the initial presentation of the word in the hard condition, participants could neither see what word they were entering, nor what they had already entered.

Because participants had to alternate between the two tasks after every response, they had to keep track of whether a carry was in progress for the subtraction task and what the word was for the text-entry task while performing the other task.

In the support condition a marker on the screen indicated whether a carry was in progress in the subtraction task. Figure 3.1 shows this condition. The '|' indicates that there is currently no carry in progress. However, as soon the previous column resulted in a carry (e.g., after a column like 3 - 4), the '|' turned into a '1'. Thus, in the support condition it was not necessary to keep track of the problem state mentally: when a '1' was shown on screen, there was a carry in the previous column.

Stimuli and Apparatus

The stimuli for the subtraction task were generated anew for each participant. The subtraction problems in the hard version always featured six carries, and resulted in 10-digit answers. The 10 letter words for the hard version of the text-entry task were handpicked from a list of high-frequency Dutch words (CELEX database; Baayen et al., 1993) to ensure that similarities between words were kept at a minimum. These stimuli were also used in the easy text-entry task, except that the letters within the words were scrambled (under the constraint that a letter never appeared twice in a row). Thus, participants were presented pseudo-random sequences of letters that they had to enter one-by-one in the easy condition. By scrambling the words, we controlled for letter-based effects, while preventing the use of strategies to predict the next letter.

The experiment was presented full screen on a 20.1" monitor. Participants were sitting at a normal viewing distance, about 70 cm from the screen. For recording pupil dilation an Eyelink 1000 table-mounted eye tracker of SR Research was used. We recorded one eye, with a sampling frequency of 500 Hz. To improve measurements, participants were seated with their heads positioned in a padded head- and chin-rest.

Procedure

Each trial started with the presentation of a calibration circle for the eye tracker. After the calibration circle a fixation cross was presented for 6 seconds, to allow pupil dilation to return to baseline. The fixation cross was followed by two horizontally aligned colored circles representing the tasks. The color of the circles indicated the difficulty levels of the tasks (on the left for the subtraction task, on the right for the text-entry task; green for easy, red for hard). The circles stayed on the screen for 1 second, followed by a fixation cross for 600 ms, after which the subtraction and text-entry tasks appeared. Participants had to begin with the subtraction task, and then alternate between the two tasks. After completing both tasks, a feedback screen was shown for 2 seconds, indicating how many letters and digits were entered correctly. Before the next trial started, a fixation screen was shown for 2 seconds.

The experiment consisted of a practice block and two experimental blocks. One of the experimental blocks contained the support condition; the order was counter-balanced over participants. The practice block consisted of 12 single task trials (4 subtraction trials with 10 columns visible, 4 subtraction trials with one column visible, and 4 text-entry trials), followed by a block of 4 multitasking trials: all combinations of subtraction and text-entry (*easy-easy*, *hard-easy*, *easy-hard*, and *hard-hard*). Both experimental blocks consisted of 28 multitasking trials. Before the second block the subtraction task was practiced again, to familiarize the participants with using the carry indicator if they did not use this in the first block, or with performing the task without the indicator in the other case. Subtraction and text-entry conditions were randomized within a block. The complete experiment consisted of 56 experimental trials, and lasted for about 90 minutes. In between blocks participants could take a short break. At the start of the multitask practice block and the two experimental blocks the eye tracker was (re) calibrated.

Results

We will discuss the results of the experiment on the basis of our experimental questions. First, we will discuss how the No-Support condition gives experimental support for a problem state bottleneck. We will then turn to the Support condition, to see if the effects of the problem state bottleneck disappear for the subtraction task when external support is provided. Finally, we will discuss mental workload.

We only analyzed the data from the experimental phase of the task. A response time in the subtraction task is defined as the time between a response in the text-entry task and a response in the subtraction task; a response time in the text-entry task as the time between a response in the subtraction task and a response in the text-entry task. First responses of each trial were removed. Outliers were removed from the data (RTs < 250 ms or $> 10{,}000$ ms), after which we removed data exceeding three standard deviations from the mean per condition per participant (in total 2.2% of the data was removed). All *F*- and *p*-values are obtained from repeated measure ANOVAs; all error bars depict standard errors. Accuracy data were transformed using an arcsine transformation before being submitted to the ANOVA. We will not discuss all effects in the text, but only the ones relevant to our questions. However, all ANOVA results are listed in Tables 3.1, 3.2, 3.3, 3.4, and 3.5.

The Problem State Bottleneck: Replication

The No-Support condition of the current experiment is a replication of Experiment 1 in Borst, Taatgen, & Van Rijn (2010), which was the first in a series of three experiments that we used to argue in favor of a problem state bottleneck. The current results replicate

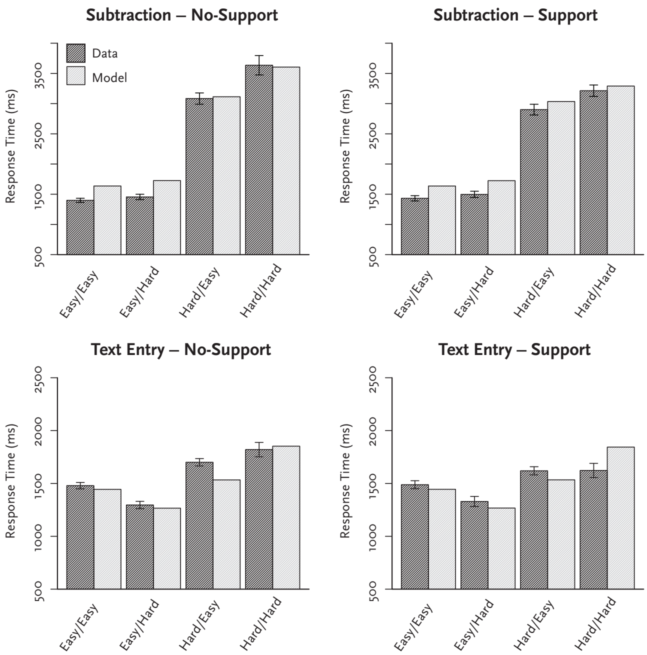


Figure 3.2 Response times. Easy/Hard means Easy Subtraction – Hard Text-Entry, etc. Error bars are standard errors.

the effects in the original data. The left panels of Figure 3.2 show the response times in the No-Support condition (ANOVA results are listed in Table 3.1). On top the response times on the subtraction task are shown. First, we see a large increase in response times with Subtraction Difficulty: when subtraction was hard, response times were much higher than when subtraction was easy. More interestingly, when both tasks were hard, there was an additional increase in response times, as shown by a significant over-additive interaction effect between Subtraction Difficulty and Text-Entry Difficulty. This interaction effect was taken as an indication of a problem state bottleneck: when participants had to maintain a problem state for both tasks, response times increased considerably as compared to when they had to maintain a problem state for only one task.

A similar effect can be seen in the response times on the text-entry task (Figure 3.2, left side, lower panel). Here, response times were lower when text-entry was hard than when text-entry was easy (we discuss this effect in the model section below). However, response times increased when subtraction was hard as well: the *hard-hard* condition. Again, because an additional problem state is required in the other task, we see an increase in response times on the current task. Statistically, this is shown by a significant interaction effect between Subtraction Difficulty and Text-Entry Difficulty.

The accuracy data of both tasks also show this effect, as shown in Figure 3.3, left panels (ANOVA results in Table 3.2). While accuracy naturally decreases with task

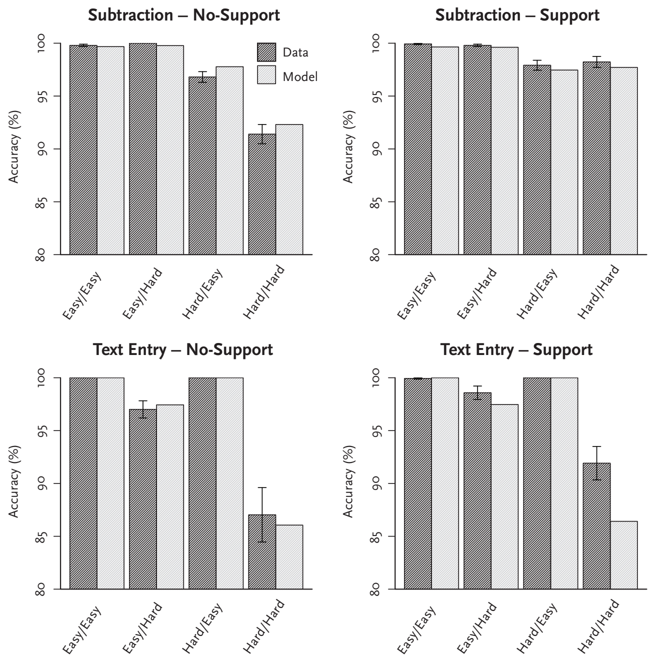


Figure 3.3 Accuracy. Easy/Hard means Easy Subtraction – Hard Text-Entry, etc. Error bars are standard errors.

difficulty of the task itself, it decreases even more when the other task is hard as well (shown by significant interaction effects between Subtraction Difficulty and Text-Entry Difficulty).

Summarizing, we see that response times increase and accuracy decreases when participants had to maintain two problem states as compared to zero or one. Previously, these interaction effects were taken as an indication of a problem state bottleneck. The question is now whether these interaction effects disappear in the subtraction task when external support is provided.

External Support: Bypassing the Bottleneck

The right panels of Figure 3.2 show the response times in the Support Condition. With respect to the response times on the subtraction task, a significant three-way interaction between Support, Subtraction Difficulty, and Text-Entry Difficulty (Table 3.3) shows that the two-way interaction between Subtraction Difficulty and Text-Entry is smaller in the Support condition than in the No-Support condition. Thus, participants were faster in the *hard-hard* condition with Support than without Support. However, as can be seen in Figure 3.2 and Table 3.1, the two-way interaction between Subtraction Difficulty and Text-Entry Difficulty was also significant with Support: even with external support

Table 3.1 ANOVA results of the response time data; separate for Support and No-Support.

	RT No-Support			RT Support		
Source	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2
			Subtraction Task			
Subtraction Difficulty	357.9	< .001	.94	531.3	< .001	.96
Text-Entry Difficulty	22.0	< .001	.49	15.0	< .001	.40
Subtraction × Text-Entry	20.0	< .001	.47	14.5	< .001	.39
			Text-Entry Task			
Subtraction Difficulty	133.6	< .001	.85	46.4	< .001	.67
Text-Entry Difficulty	< 1	–	–	2.6	.12	.10
Subtraction × Text-Entry	26.5	< .001	.53	8.6	.007	.27

RT = response times.

Table 3.2 ANOVA results of the accuracy data; separate for Support and No-Support. Note that for the accuracy of the text-entry task we did not find any effects involving Support, which is why we collapsed over Support.

	Acc No-Support			Acc Support		
Source	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2
			Subtraction Task			
Subtraction Difficulty	80.4	< .001	.77	36.8	< .001	.62
Text-Entry Difficulty	45.0	< .001	.66	< 1	-	-
Subtraction × Text-Entry	58.2	< .001	.72	1.1	.3	.05
			Text-Entry Task			
Subtraction Difficulty	25.9	< .001	.53			
Text-Entry Difficulty	173.0	< .001	.88		*Same as No-Support*	
Subtraction × Text-Entry	28.1	< .001	.55			

Acc = accuracy.

participants show an increase in response times in the *hard-hard* condition. Thus, the effect of the problem state bottleneck decreases, but does not fully disappear.

With respect to the response times of the text-entry task, we also observed a significant three-way interaction effect of Support, Subtraction Difficulty, and Text-Entry Difficulty. When external support was provided for the subtraction task, the effects of the problem state bottleneck also decreased in the text-entry task: participants were faster in the *hard-hard* condition. However, also here the two-way interaction effect between Subtraction Difficulty and Text-Entry Difficulty is still present with external support.

The right panels of Figure 3.3 show the accuracy data in the Support condition. Here we see that the two-way interaction effect completely disappears for the subtraction task (the three way interaction is significant, Table 3.3). Thus, participants no longer show the decrease in accuracy in the *hard-hard* condition: the effect of the problem state bottleneck disappeared. In the text-entry task there was no difference between the

Table 3.3 Overall ANOVA results, on the left for response times, on the right for accuracy.

	Response Times			Accuracy		
Source	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2
			Subtraction Task			
Support	8.02	.009	.26	65.7	< .001	.74
Subtraction	484.2	< .001	.95	103.8	< .001	.82
Text-Entry	26.3	< .001	.53	18.78	< .001	.45
Support × Subtraction	27.6	< .001	.55	66.8	< .001	.74
Support × Text-Entry	4.22	.05	.15	12.7	.002	.36
Subtraction × Text-Entry	29.35	< .001	.56	24.7	< .001	.52
Support × Sub. × Text-Entry	5.05	.03	.18	21.4	< .001	.48
			Text-Entry Task			
Support	2.85	.10	.11	3.20	.09	.12
Subtraction	105.5	< .001	.82	25.7	< .001	.53
Text-Entry	1.17	.29	.05	149.4	< .001	.87
Support × Subtraction	32.7	< .001	.59	< 1	–	–
Support × Text-Entry	3.96	.06	.15	3.87	.06	.14
Subtraction × Text-Entry	19.7	< .001	.46	27.1	< .001	.54
Support × Sub. × Text-Entry	9.29	.006	.29	< 1	–	–

Subtraction = Subtraction Difficulty, Text-Entry = Text-Entry Difficulty.

Support and the No-Support conditions: in both conditions they make most mistakes in the *hard-hard* condition.

In summary, in the response times of the subtraction task the effect of the problem state bottleneck decreased, but did not disappear. In the accuracy data, on the other hand, performance in the subtraction task reached no-bottleneck levels with support. This indicates that external support can indeed help bypassing the problem state bottleneck, but does not bring performance fully back to normal levels. Below we will discuss how our model accounts for this. However, we will first describe the mental workload results.

Pupil Dilation: Mental Workload

Measuring pupil dilation served two goals: (1) investigating whether the problem state bottleneck leads to an increase in mental effort; and (2) seeing if the level of mental effort changes in the support condition. We calculated percentage change in pupil dilation as compared to the average dilation during the fixation screen before each trial; only data of stable fixations were taken into account. For each step in a trial (entering a digit or letter) the maximum pupil dilation was taken, which was then averaged per condition and participant. The results are shown in Figure 3.4, the ANOVA results reported in Table 3.4 (all conditions) and Table 3.5 (collapsed over Support). As we did

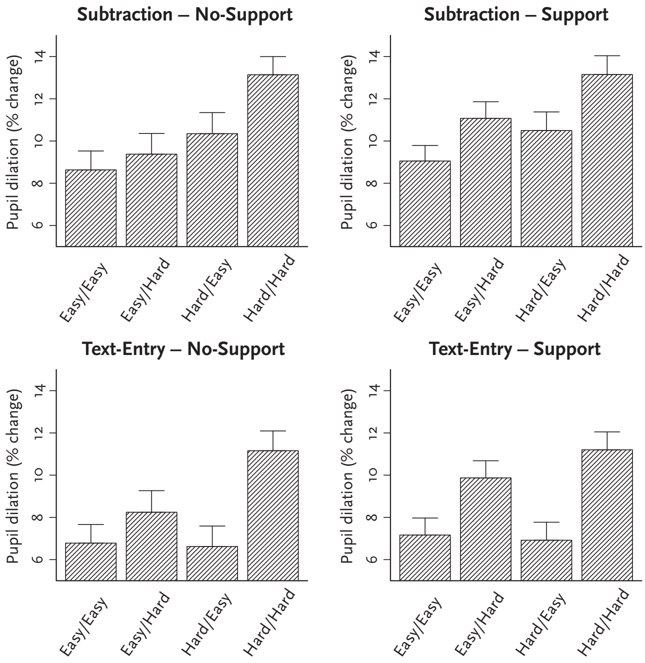


Figure 3.4 Pupil dilation results. Easy/Hard means Easy Subtraction – Hard Text-Entry, etc. Error bars are standard errors.

not find a difference in pupil dilation between the support and the no-support condition for either task (no effects involving Support were significant, Table 3.4), we collapsed over Support for both tasks.

The top row of Figure 3.4 shows pupil dilation in the subtraction task. While it seems as if the interaction effect between Subtraction Difficulty and Text-Entry Difficulty is somewhat smaller with external support, this effect did not reach significance. Overall, the two-way interaction between Subtraction Difficulty and Text-Entry Difficulty was significant (see Table 3.5), as were the main effects of Subtraction Difficulty and Text-Entry Difficulty. Thus, pupil dilation increased with task difficulty, and increased even more when both tasks were hard. This replicates the effects that were found in the response time and accuracy data.

The bottom row of Figure 3.4 shows pupil dilation during the text-entry task. The overall interaction effect between Subtraction Difficulty and Text-Entry Difficulty was even more pronounced for the text-entry task than for subtraction: highest pupil dilation levels were observed in the *hard-hard* condition (Figure 3.4). However, again we did not find a significant difference between the support and the no-support condition. In addition to the two-way interaction effect, the main effect of Text-Entry Difficulty was also significant. Thus, in the text-entry task pupil dilation increased considerably when text-entry was hard, but not when subtraction was hard. However, when both tasks were hard an additional increase in pupil dilation was observed.

Table 3.4 Overall ANOVA results for the pupil dilation data.

	Pupil Dilation		
Source	$F(1,23)$	p	η_p^2
	Subtraction Task		
Support	1.86	.19	.07
Subtraction	15.73	< .001	.41
Text-Entry	25.13	< .001	.52
Support × Subtraction	1.24	.28	.05
Support × Text-Entry	< 1	–	–
Subtraction × Text-Entry	7.25	.01	.24
Support × Sub. × Text-Entry	1.47	.24	.06
	Text-Entry Task		
Support	2.98	.10	.11
Subtraction	3.51	.08	.13
Text-Entry	56.14	< .001	.71
Support × Subtraction	1.05	.32	.04
Support × Text-Entry	< 1	–	–
Subtraction × Text-Entry	24.05	< .001	.51
Support × Sub. × Text-Entry	2.16	.15	.09

Subtraction = Subtraction Difficulty, Text-Entry = Text-Entry Difficulty.

Table 3.5 Overall ANOVA results for the pupil dilation data.

	Pupil Dilation		
Source	$F(1,23)$	p	η_p^2
	Subtraction Task		
Subtraction Difficulty	15.79	< .001	.41
Text-Entry Difficulty	24.76	< .001	.52
Subtraction × Text-Entry	7.33	.01	.24
	Text-Entry Task		
Subtraction Difficulty	3.56	.07	.13
Text-Entry Difficulty	55.86	< .001	.71
Subtraction × Text-Entry	24.32	< .001	.51

In summary, we found strong interaction effects between Subtraction Difficulty and Text-Entry Difficulty. These indicate that the performance decrease caused by the problem state bottleneck is also reflected in an increase in mental workload. When external support was provided, this increase in mental workload did not disappear for either task.

Cognitive Model

To account for the observed data, we adapted our computational model of the problem state bottleneck (Borst, Taatgen, Stocco, & Van Rijn, 2010; Borst, Taatgen, & Van Rijn, 2010). This model was developed in the ACT-R cognitive architecture (e.g., Anderson, 2007), and uses threaded cognition theory to account for the multitasking aspects of the task (Salvucci & Taatgen, 2008, 2011). Using a cognitive architecture ensures that the components of the model have been validated earlier, which makes it meaningful to take for instance the memory, visual, and motor components of the task into account (e.g., Cooper, 2007; Newell, 1990). We do not describe the complete model here, but refer the interested reader instead to Borst, Taatgen, & Van Rijn (2010).

For the current paper, the problem state component is our main interest. The assumption of the model is that the problem state resource can only maintain one chunk of information at a time. Thus, as long as at most one of the tasks is hard, the model can do the task without a problem – because then at most one problem state is required – but when both tasks are hard the model can only maintain one problem state, which results in interference. The model assumes that in the *hard-hard* condition, on each step in a trial the problem state resource is swapped out. That is, problem state information of the now current task is restored to the problem state resource, while problem state information of the previous task is moved to declarative memory. Thus, when the model switches to the other task, it first retrieves the necessary problem state information from declarative memory, restores this to the problem state resource, and only then performs the task. This takes time (a memory retrieval and 200 ms problem state restoration costs; Anderson, 2005), which results in increased response times in the *hard-hard* condition. Furthermore, incorrect problem states are sometimes retrieved from memory, resulting in lower accuracy scores in the *hard-hard* condition.

The grey bars in the left panels of Figure 3.2 and 3.3 show the fit of this model to the original task[1]. As can be observed, the model accounts well for the interaction effects in both response times and accuracy data, and also matches quite well to the absolute response times and accuracy data of the task (R^2- and RMSD-values are reported in Table 3.6). For instance, while we did not add this explicitly to the model, response times are lower in the hard text-entry condition than in the easy text-entry condition. This is caused by the fact that in the easy condition the model has to read which letter it has to enter before it can search for a button and click on it, while in the hard condition the model (and participants) already knows what word it is entering. This saves visual perception time, and thus results in lower response times in the hard text-entry condition.

We extended the model to also perform the subtraction task in the support condition. There were two basic options: either the model always uses the support indicator on the screen, which would result in equal response times in the *hard subtraction – easy text-entry* and the *hard-hard* condition, or it only uses the indicator when it cannot use its problem state, in the *hard-hard* condition. The latter option seems to be the most rational one: using the problem state to remember whether a carry is in progress

[1] *We fit the model to the data in the no-support condition by estimating retrieval times and retrieval errors from declarative memory, and mouse- and eye movements. The model code is available from http://www.jelmerborst.nl/models/.*

Table 3.6 **Model fit.**

	Measurement	R^2	RMSD
RT	Subtraction No-Support	1.0	181 ms
	Subtraction Support	1.0	171 ms
	Text-Entry No-Support	0.88	88 ms
	Text-Entry Support	0.72	124 ms
Acc	Subtraction No-Support	0.99	.68 %
	Subtraction Support	1.0	.38 %
	Text-Entry No-Support	1.0	.53 %
	Text-Entry Support	1.0	2.8 %

RT = Response Times, Acc = accuracy, RMSD = root mean squared deviation.

takes less time than having to look at the indicator on each step of a trial (cf. the lower response times in the hard text-entry condition in the previous paragraph). Thus in total it will take less time to do the task when the support indicator is only used in the *hard-hard* condition. When we implemented this strategy in the model, that is, using the problem state resource in the *hard-easy* condition, but the support indicator in the *hard-hard* condition, this led to a good model-data fit (see Figure 3.2 and 3.3; please note that we did not make any additional changes to the model, all parameters were kept at the same values as for the no-support condition). On the one hand, implementing this strategy resulted in a small interaction effect in response times in the support condition (in the *hard-hard* condition the support indicator has to be processed, this takes more time than doing the task mentally in the *hard-easy* condition). On the other hand, it also results in a complete absence of the interaction effect in the accuracy data (as the model does not make mistakes anymore because of retrieving incorrect problem states). Thus, it seems that participants use the externally presented support only when it helps them to do the task faster than a mental strategy would allow.

It should be noted that the model always uses its problem state resource to *process carries* in the subtraction task, also in the *hard-hard* condition *with* support. Thus, when it has to process a carry, it will use its problem state to represent the intermediate solution (e.g., when solving '6 - 4' with a carry, it will use the problem state to represent '5 - 4'). This is why the model predicts no changes to the interaction effects for the text-entry task when external support is presented: It always has to retrieve the text-entry problem state from declarative memory and restore it to the problem state resource before it can start the text-entry task. However, the data show a small decrease of the interaction effects in the support condition in the text-entry task. A simple explanation could be that participants do not need to overwrite their text-entry problem state when using the support indicator for the subtraction task. This should lead to a complete absence of the interaction effect though, both in response times and accuracy. While we see a decrease, the interaction effect is still present. As we have no strong hypothesis about what happens, we decided against making post-hoc changes to the model to fit this.

Summarizing, the model accounts well for the main effects in the data. While we might have expected to find no interaction effects at all in the support condition for the subtraction task, the model shows that rational behavior does lead to an interaction in the support condition, albeit a smaller one than in the no-support condition. Thus, it seems that it is possible to use problem state information in the environment, but that people only do so when an environment-based strategy is faster than a mental strategy.

General Discussion

In this article we investigated a major bottleneck in human multitasking: the problem state (Borst, Taatgen, & Van Rijn, 2010). It was previously shown that this bottleneck can have considerable influence on performance, both in the lab and in more real-world tasks (Salvucci & Bogunovich, 2010). We therefore looked at how we can design tasks in such a way that the problem state bottleneck can be bypassed. In a dual-task experiment, it was shown that if problem state information is presented on the screen, the negative effects of the problem state bottleneck diminish. Accuracy levels came back to non-bottleneck levels, while response times improved, but did not reach non-bottleneck levels. The computational model showed that this is rational behavior: Participants performed the task as fast as possible, which in the single problem state case meant that they did it mentally, while they used the external support when a problem state was required for both tasks. These results seem to indicate that the problem state bottleneck can be avoided by presenting information in the environment, but that users will only use this information when it leads to faster and less error-prone behavior. Furthermore, the model showed that participants still use their problem state resource for subtraction in the support condition, because otherwise the interaction effects in the text-entry task should have disappeared.

It is not surprising that presenting external support improves performance on a task; the beneficial effects of offloading mental representations to the environment have been described before (e.g., Hollan, Hutchins, & Kirsh, 2000; Kirsh, 1995; Wickens, 1992). However, the current experiment shows that presenting information in the environment only helps in certain cases. Using the model, it is possible to predict exactly when external support is helpful, and when not. In general, we can conclude that it only helps to present external support when users need more than one problem state to carry out a task. While in other cases it might still be used as a memory aid, there are limits on presenting information in the environment. The current research indicates that it is often not necessary to present external information, and that it will not even be used when a mental strategy is faster. Moreover, while the current very simple interface already shows that the costs of visually processing external support have an influence on the task, this is much more important with a real-life interface. When multiple sources of external support are present (as is often the case in real-life systems), this will increase the costs of actually using it, making it important to only present external support when it improves performance.

As we just described, the current research shows that external support is only helpful when multiple problem states are required for performing a task. However, on the other hand it indicates that already with two concurrent problem states it can be profitable to present information externally. This runs counter to classical ideas that we only have to offload internal representations if we cross a threshold of about 5 items. For instance, based on the classical idea of a working memory capacity of 7 ± 2 items (Miller, 1956), Wickens states that "The 7 ± 2 limit is a critically important one in system design." (1992, p. 222). Based on the current research we would argue for a much lower limit of only one item. However, this is also dependent on the costs of processing the external support: naturally, it is only helpful to present support when the gains are higher than the costs.

Besides behavioral measurements, we also recorded pupil dilation during the experiment. Pupil dilation is assumed to reflect mental effort in a task (e.g., Beatty, 1982; Steinhauer & Hakerem, 1992). Where we previously reported interaction effects in response times and accuracy, we now show that the problem state bottleneck also leads to an over-additive increase in mental effort: we observed higher dilation in the *hard-hard* condition than would be expected based on the separate hard conditions. This is not simply a reflection of increased response times: in the *easy subtraction – hard text-entry* condition we see for example faster response times (Figure 3.2), but higher pupil dilation than in the *easy-easy* condition (Figure 3.4). Interestingly, we did not find a significant difference in mental workload between the support and the no-support condition. This could indicate that while participants offload problem state processing to the screen, it still leads to an increase in mental effort (at least as indicated by pupil dilation). However, as there seems to be a slight difference between the conditions (Figure 3.4), additional experiments are necessary to make this clear.

Given that an objective measure of mental workload, pupil dilation, does not show a difference between the support and no-support conditions, it is likely that also a subjective measure of workload (i.e. questionnaires) would not yield a difference. However, we see in the behavioral data that performance does improve significantly when external support is provided. This indicates that asking users if a certain task environment is a useful improvement is not sufficient, but that detailed measures like response times and low-level errors have to be taken into account when designing user interfaces. Using these measures, cognitive models can then help in identifying bottlenecks in behavior, and how these can be bypassed(see also Gray, 2008, on the use of cognitive architectures in human factors). As shown, a model can for instance be used to predict when external support will be useful to the users of the task, and when the visual processing costs are too high for it to be useful.

We conclude that it is possible to bypass the behavioral effects of the problem state bottleneck by presenting external support. However, when giving external support, it should be taken into account that it is only useful when users need more than one problem state to perform a task, and when the processing costs of the support are not higher than the behavioral gains.

The Neural Correlates of Problem States: Regions-of-Interest Analysis

In which we investigate the neural correlates of problem states by testing a priori fMRI predictions of our cognitive model with a regions-of-interest analysis.

This chapter was previously published as:

Borst, J.P., Taatgen, N.A., Stocco, A. & Van Rijn, H. (2010). The Neural Correlates of Problem States: Testing fMRI Predictions of a Computation Model of Multitasking. *PLoS ONE, 5*(9), e12966.

4

Chapter

Chapter 4

The Neural Correlates of Problem States: Regions-of-Interest Analysis

Abstract

Background: *It has been shown that people can only maintain one problem state, or intermediate mental representation, at a time. When more than one problem state is required, for example in multitasking, performance decreases considerably. This effect has been explained in terms of a problem state bottleneck.*

Methodology: *In the current study we use the complimentary methodologies of computational cognitive modeling and neuroimaging to investigate the neural correlates of this problem state bottleneck. In particular, an existing computational cognitive model was used to generate a priori fMRI predictions for a multitasking experiment in which the problem state bottleneck plays a major role. Hemodynamic responses were predicted for five brain regions, corresponding to five cognitive resources in the model. Most importantly, we predicted the intraparietal sulcus to show a strong effect of problem state manipulations.*

Conclusions: *Some of the predictions were confirmed by a subsequent fMRI experiment, while others were not matched by the data. The experiment supported the hypothesis that the problem state bottleneck is a plausible cause of the interference in the experiment and that it could be located in the intraparietal sulcus.*

Introduction

One of the challenges for research on multitasking is to explain why some tasks can be performed together without a problem (e.g., talking and walking), while other tasks clearly interfere with each other (e.g., talking and reading). According to so-called multiple-resource theories, interference occurs when multiple tasks require the same cognitive or peripheral resources (e.g., Navon & Gopher, 1979; Salvucci & Taatgen, 2008; Wickens, 2002). An obvious example is our visual system: we can only look at one thing at a time. There is empirical evidence that the same principle might hold for cognitive resources: for instance indicating that we can only retrieve one fact at a time from declarative memory (e.g., Anderson & Lebiere, 1998). The impact of a concurrent request to a particular resource depends on the time scale of multitasking: whether it is truly concurrent multitasking (e.g., driving and calling), or whether the task can be characterized as 'sequential multitasking' (e.g., writing a paper and answering the phone; Salvucci, Taatgen, et al., 2009).

A resource that causes considerable interference in both concurrent and sequential multitasking is the problem state resource. This resource is used for maintaining intermediate task representations. For instance, when mentally solving the algebra problem $3x - 10 = 2$ it is used to store $3x = 12$ (e.g., Anderson, 2005). In a series of experiments we have shown that the problem state resource acts as a bottleneck in sequential multitasking (Borst, Taatgen, & Van Rijn, 2010). When multiple tasks needed to store intermediate results, interference was observed. However, when only one of the tasks required access to intermediate results no interference was found. To account for these experimental results we developed a computational cognitive model that showed that a 'problem state bottleneck' could explain the behavioral data.

The goal of this paper is to explore the neural underpinnings of the problem state bottleneck and to further validate our cognitive model. To these ends, the model was used to generate *a priori* predictions of hemodynamic activation patterns in five predefined brain areas for a triple-task. Subsequently, an fMRI experiment was conducted, and the model predictions were compared to the data. Some of the predictions were confirmed while others did not match with the data. In general the results corroborate the model and provide further evidence (see e.g., Anderson, 2005) that the intraparietal sulcus is a probably location for the problem state resource. In the remainder of this paper we will first introduce the theory related to the problem state bottleneck, followed by a description of the experiment, the model, and the fMRI predictions. Finally, we will discuss the correspondence between the predictions and the fMRI data, and the implications for the problem state bottleneck hypothesis.

The Problem State Bottleneck

The problem state resource is the part of working memory responsible for storing intermediate representations in a task. For instance, the problem state can be used to store an intermediate state of an algebra problem, as mentioned above. An everyday example is asking for driving directions, during which one needs the problem state

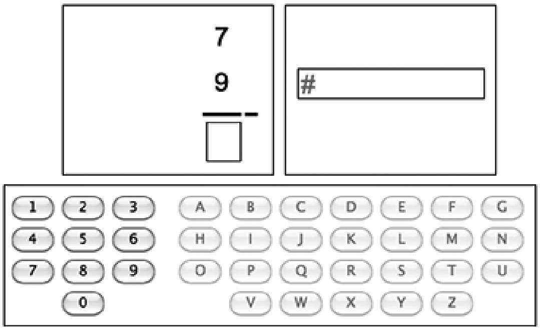


Figure 4.1 Interface of the experiment. Note that the disabled task is masked by #-marks.

resource to store at which street one should turn to arrive at the destination. Note that if the same information is present in the world, that is, if one works out the algebra problem on paper or follows road signs to the destination, it is not necessary to maintain a problem state. An important functional characteristic of the problem state resource is that its contents are directly accessible for the task at hand. This in contrast to other elements in working memory, which are only available at a time cost (e.g., McElree, 2001).

The concept of a central problem state resource originates from a series of neuro-imaging experiments by Anderson and colleagues, who found that the Blood-Oxygen Level-Dependent (BOLD) signal in the posterior parietal cortex correlates with the number of transformations of mental representations (e.g., Anderson, 2005; Anderson, Albert, et al., 2005; Anderson et al., 2003; Sohn et al., 2005).

Previously, we have conducted a number of experiments investigating the nature of this resource (Borst & Taatgen, 2007; Borst, Taatgen, & Van Rijn, 2010). These experiments showed that people can only maintain one problem state at a time. When a problem state was required for more than one task, performance decreased considerably, indicating a processing bottleneck. To account for these results we constructed a cognitive model based on the threaded cognition theory (Salvucci & Taatgen, 2008) and the cognitive architecture ACT-R (Anderson, 2007). The model fit well to the data (see the next section), further corroborating the hypothesis of a problem state bottleneck as a plausible explanation of multitasking interference. The next section will discuss how the model was used to generate fMRI predictions for the current study.

A Priori Model Predictions

To validate cognitive models, it is common practice to compare model data to behavioral data. For instance, if response times and accuracy scores correspond well between model and data, it is assumed that a model gives a plausible explanation of the data. However, many cognitive models have a complexity that cannot be accounted for by using only behavioral measurements (e.g., Myung, 2000; Roberts & Pashler, 2000). One solution is to use predictions: first use a cognitive model to predict the outcome of an experiment, and only conduct the experiment afterwards (see for examples Salvucci & Macuga, 2002; Taatgen & Anderson, 2008; Taatgen et al., 2007). Nevertheless, there are so many degrees of freedom in developing a model that models are often under-constrained by behavioral data. To increase the constraints on models that are developed in the cognitive architecture ACT-R, a methodology was developed for mapping model activity on brain activity (for a concise explanation, see Anderson et al., 2008). This way, models are not only constrained by behavioral data, but also by neuroimaging data. The next sections will describe how this methodology was used to generate *a priori* neuroimaging predictions from our model. We will first describe the experimental setup and the model itself, followed by the actual predictions.

The triple task

The task for which we generate BOLD-predictions is a triple task in which participants have to perform a subtraction task, a text-entry task, and a listening comprehension task (similar to Experiment 3 in Borst, Taatgen, & Van Rijn, 2010). The subtraction and text-entry tasks both have an easy version for which maintaining a problem state is not required to perform the task, and a hard version for which maintaining a problem state is required to perform the task correctly. In half of the trials, participants also had to listen to a short story on which they were quizzed after the trial. To measure baseline performance on the listening task, we included an 'Only Listening' condition in which participants only had to do the listening task. Thus, the experiment has a $2 \times 2 \times 2 + 1$ design (Subtraction Difficulty (easy/hard) × Text-Entry Difficulty (easy/hard) × Listening (yes/no) + Only Listening).

Figure 4.1 shows the graphical interface of the experiment. The subtraction and text-entry tasks were presented at the same time on two different panels of the interface; participants had to alternate between these tasks. After entering a digit in the subtraction task, the subtraction panel was disabled, forcing the participant to subsequently enter a letter. After entering a letter, the text-entry panel was disabled and the subtraction panel became available again. In half of the trials, the listening task had to be performed at the same time as the other two tasks. Thus, this paradigm allows us to study both concurrent (listening and subtraction/text-entry), and sequential multitasking (alternating between subtraction and text-entry).

The interface for the subtraction task is shown in the left panel of Figure 4.1. In the subtraction task participants had to solve multi-column subtraction problems in standard right-to-left order. However, at each point in time, only one column was visible. Although the problems were presented column by column, the participants were trained to perceive the separate columns in a trial as one 10-column subtraction problem (in the practice phase participants started out with a normal 10-column layout, only later they switched to solving the problems column by column). Participants had to enter the digits by clicking on the on-screen keypad with the mouse. In the easy, no problem state version, the upper digit was always larger or equal to the lower one;

these problems could be solved without 'borrowing'. In contrast, the hard version required participants to borrow six times out of 10 possible columns. The assumption, supported by the results of Borst, Taatgen, and Van Rijn (2010), is that participants have to use their problem state resource to keep track of whether a 'borrowing' is in progress.

The interface for the text-entry task is shown on the right in Figure 4.1. Participants had to enter 10-letter strings by clicking on the on-screen keypad. In the easy version these strings were presented one letter at a time and participants had to click the corresponding button on the keypad. In the hard version, a 10-letter word was presented once at the start of a trial. Once a participant clicked on the first letter, the word disappeared and the remaining letters had to be entered one at a time, without feedback. Thus, after the initial presentation of the string in the hard condition, participants could neither see what word they were entering, nor what they had already entered. Results by Borst, Taatgen, and Van Rijn (2010) provide evidence that participants use their problem state resource to keep track of the process.

The listening comprehension task had to be performed during half of the trials. This task consisted of listening to a short story about which a multiple-choice question would be asked at the end of the trial. After answering the question, participants received accuracy feedback. According to existing models of language processing in ACT-R, this task does not require maintenance of problem states, but draws on different cognitive resources (Lewis & Vasishth, 2005; Lewis et al., 2006). Furthermore, the listening task did not affect the problem state-related outcomes of Experiment 3 in Borst, Taatgen, and Van Rijn (2010), also indicating an absence of problem state usage. This, in turn, indicates that problem state interference does not depend on the number of tasks, but on the particular cognitive resources used by the tasks. In the 'only listening' condition a fixation cross was shown instead of the subtraction and text-entry tasks.

Because participants had to alternate between the subtraction and text-entry tasks after every letter and digit, they had to maintain intermediate state information for the other task (when it was hard) while giving a response on the current task. Based on the threaded cognition theory (Salvucci & Taatgen, 2008), we predicted that it is not possible to maintain more than one problem state at a time, and therefore expected to find interference when participants have to use a problem state for both tasks. As the listening task was assumed not to use the problem state resource, it was expected that problem state interference was independent of the listening task.

The results of the behavioral experiment of Borst, Taatgen, and Van Rijn (2010) were as follows. Response times were considerably higher and accuracy lower in the *hard subtraction – hard text-entry* condition than in the other conditions. In fact, we found an interaction effect of Subtraction Difficulty and Text-Entry Difficulty both in response times and accuracy. The listening task had little behavioral effect; it was limited to a small increase in response times in the subtraction task when the listening task was added. Because the subtraction and text-entry tasks were performed sequentially, it is unlikely that the observed interaction was caused by condition-specific differences between the easy and hard conditions: only problem states had to be maintained while doing the other task (see for a much more elaborate discussion of these results Borst, Taatgen, & Van Rijn, 2010, in particular Experiment 2). Thus, in line with the problem state bottleneck hypothesis, the strongest interference occurred in the *hard subtraction – hard text-entry* condition, indicating that participants could not maintain two problem states at the same time.

The cognitive model

To account for these results, a model was developed in the cognitive architecture ACT-R, using the threaded cognition theory to handle multitasking. First we will introduce ACT-R and threaded cognition, followed by a description of the model itself.

The cognitive architecture ACT-R (Anderson, 2007) describes human cognition as a set of independent modules – cognitive resources – that interact through a central production system. For instance, it uses visual and aural modules for perception and a motor module to interact with the world. Besides these peripheral modules ACT-R also has a number of central cognitive modules: the procedural module that implements the central production system, the declarative memory module, the goal module, and the problem state module. All modules operate in parallel, but each module in itself can only proceed serially (Byrne & Anderson, 2001). Thus, the visual module can only perceive one object at a time and the memory module can only retrieve one fact at a time.

Threaded cognition (Salvucci & Taatgen, 2008, 2011; Salvucci, Taatgen, et al., 2009) extends ACT-R by allowing multiple tasks – called threads – to be active at the same time. However, because the cognitive resources are serial in nature, the key assumption of threaded cognition is that although several tasks can be active at the same time, a particular resource can only be used by a single task at a time, and thus acts as a bottleneck when required by multiple tasks concurrently.

Of particular importance for the tasks at hand is ACT-R's problem state module. Although this module can hold a problem state that is accessible at no time cost, changing or restoring a problem state has been estimated to take a relatively long time (a value of 200 ms has provided a good fit in previous ACT-R models, and has been left unchanged in our models; e.g., Anderson, Qin, Stenger, & Carter, 2004; Taatgen et al., 2009). Because the problem state module can only hold one chunk of information, the module's contents have to be swapped when multiple problem states are required. When a problem state is replaced, the previous problem state remains available in long-term memory, and it can be recalled when required. However, as both retrieving an old problem state from declarative memory and updating the problem state takes time, using multiple problem states causes considerable interference. An additional effect of swapping problem states is that because older problem states need to be retrieved from memory, it is possible to retrieve an incorrect problem state from memory, resulting in behavioral errors.

The model for the triple task consists of three independent threads, one for the subtraction task, one for the text-entry task, and one for the listening task. The subtraction and text-entry threads use the visual module to perceive the stimuli and the manual module to operate the mouse. In the easy condition of the subtraction task, the model perceives the digits, retrieves a fact from memory (e.g., 5 - 2 = 3) and clicks on the corresponding button. The procedure is the same in the hard condition, up to the point when borrowing becomes necessary. When the model retrieves a fact from memory and notices that the outcome is negative (e.g., 3 - 6 = -3), the model will add 10 to the upper term, retrieve a new fact (13 - 6 = 7), and store in its problem state that a 'borrowing' is in progress. The model will then check the problem state every time the subtraction task is resumed. If a 'borrowing' is in progress, the model first subtracts 1 from the upper term before the initial retrieval is made.

In the easy version of the text-entry task, the model perceives the letter and clicks on the corresponding button. In the hard version, the model has to know the target word and the current position within that word. This information is stored in the problem state resource (e.g. "university, 4th letter"). At each step, the model uses this information to determine the next letter. To simulate the spelling processes, we implemented an additional declarative retrieval that links the current position to the next letter. Although this is a very simplified implementation of the spelling process, it was not necessary to model this aspect of the task in more details since no effects of spelling difficulty are to be expected on the problem state. After the model has determined the next letter, it clicks the appropriate button and updates the problem state to reflect that it is one position further in the word.

The listening task was modeled as a third thread. This thread aurally perceives words, retrieves lexical information related to the auditory input from memory, and builds syntactic trees. The same approach was used by Salvucci and Taatgen (2008) to model the classical reading and dictation study of Spelke, Hirst, and Neisser (1976) and by Van Rij, Van Rijn, and Hendriks (2010) and Hendriks, Van Rijn, and Valkenier (2007) to account for developmental patterns in children's ability to process pronouns. For each incoming word in the auditory module, four processing steps are taken, and two facts are retrieved from memory. This results in about 320 ms processing time per word, fast enough to keep up with the average speaking rate of 359 ms/word in the presented texts (note that the model is capable of listening to speech faster than 320 ms/word, because the auditory module can already start perceiving a word while other cognitive modules are processing the previous word). The process of answering the multiple-choice questions was not modeled, because modeling the comprehension of a question would have required linguistic processing capabilities at a level of complexity that is beyond the scope of the model. However, the model visually parses the questions when they appear on the screen.

The model explains the interference effect in the following way. In the *hard – hard* condition a problem state is needed for both the subtraction and the text-entry task. This means that the contents of the problem state resource have to be replaced on each step in a trial, increasing response times considerably. Because this is only necessary in the *hard – hard* condition, the model predicts an over-additive effect of task difficulty on response times. The number of errors will also increase with task difficulty, because older and incorrect problem states are sometimes retrieved. As the model does not use the problem state resource for the listening task, no influence of the listening task on problem state interference is predicted.

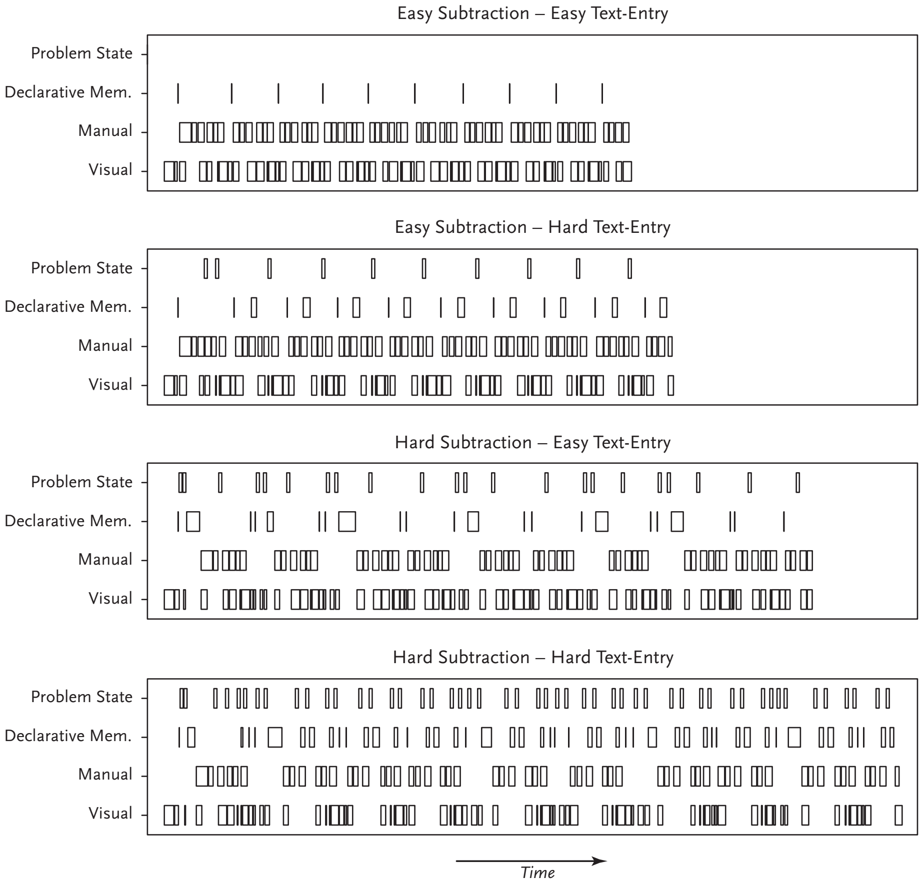


Figure 4.2 Cognitive resource usage of the model for four trial types. Time goes from left to right; boxes indicate activity of a cognitive resource. Note that only trials are depicted without the listening task.

Figure 4.2 shows how the model uses cognitive resources over the course of a trial (that is, entering 10 digits and 10 letters). The four panels show four different trial types, ranging from *easy subtraction – easy text-entry* at the top to *hard – hard* at the bottom (all without the listening task). Boxes indicate that a cognitive resource is in use. A first observation is that the length of the model traces increases with task difficulty: response times increase when the tasks get more difficult. Second, the use of the problem state resource and declarative memory also increases with task difficulty, with an over-additive increase in the *hard – hard* condition because of the problem state bottleneck. Finally, the use of the manual and visual resources is more or less constant over the different trial types, but gets more spread out in the more difficult conditions. That is, participants have to make the same number of responses in each condition, but because response times are higher these responses are spaced further apart.

Table 4.1 **ACT-R modules and associated brain regions.**

ACT-R Module	Brain Region (left hemisphere)	Size (voxels)	Talairach-Tournoux Coordinates	MNI Coordinates
Aural	Sec. auditory cortex (BA 21/22/42)	5 × 5 × 5	-45, -22, 9	-48, -21, 7
Manual	Precentral gyrus (BA 3)	5 × 5 × 4	-42, -20, 50	-42, -23, 54
Visual	Fusiform gyrus (BA 37)	5 × 5 × 4	-41, -61, -9	-43, -60, -16
Problem State	Intraparietal sulcus (BA 7/39/40)	5 × 5 × 4	-24, -63, 40	-24, -67, 44
Declarative Memory	Inferior frontal sulcus (BA 45/46)	5 × 5 × 4	-42, 23, 24	-43, 24, 25

Voxels are 3 × 3 × 3 mm. MNI = Montreal Neurological Institute. BA = Brodmann Area.

The model fit well to the behavioral data from Borst, Taatgen, and Van Rijn (2010): it fit both the interaction effects in the response times (average R^2 of .99) and in the accuracy data (average R^2 of .95; for details see Borst, Taatgen, & Van Rijn). The same model was used previously to account for the data of two other experiments (Borst, Taatgen, & Van Rijn, 2010), corroborating the model's explanation of the data. In the next section we will describe how we used the same model to generate fMRI predictions for the current experiment.

The fMRI predictions

As mentioned above, the cognitive architecture ACT-R can predict fMRI data, or to be more precise, the BOLD response (e.g., Anderson, 2007; Anderson et al., 2008). The modules of ACT-R have been mapped onto specific regions in the brain (see Table 4.1), and are assumed to predict activation in that region. The most important modules and associated brain regions for the current model are listed in Table 4.1.

ACT-R's modules are not constantly in use during the execution of a model, but operate for short periods of time (in the order of hundreds of ms). The assumption is that when a module is active the BOLD response increases in the associated brain region. The BOLD response of a certain event is modeled by a gamma function, as is customary in fMRI research (e.g., J. D. Cohen et al., 1997; Friston, 2003; Glover, 1999):

$$H(t) = m\left(\frac{t}{s}\right)^{a} e^{-\left(\frac{t}{s}\right)}$$

where t is the age of the event, m determines the magnitude of the BOLD curve, s the time scale, and a the shape. If $D(t)$ is a 0-1 demand function that indicates whether a module is active at time t, the BOLD activation at time t can be calculated by convolving $D(t)$ with the gamma function:

$$B(t) = D(t) \otimes H(t) = \int_0^t D(\tau)\, H(t-\tau)\, d\tau$$

Because the predictions were made before the experiment was run, the gamma function parameters were not fit to data but were set to default ACT-R values ($s = .75$, $a = 6$). The scaling parameter (m) was left at 1 (note that therefore only the shape of the predictions is of interest, not the magnitude).

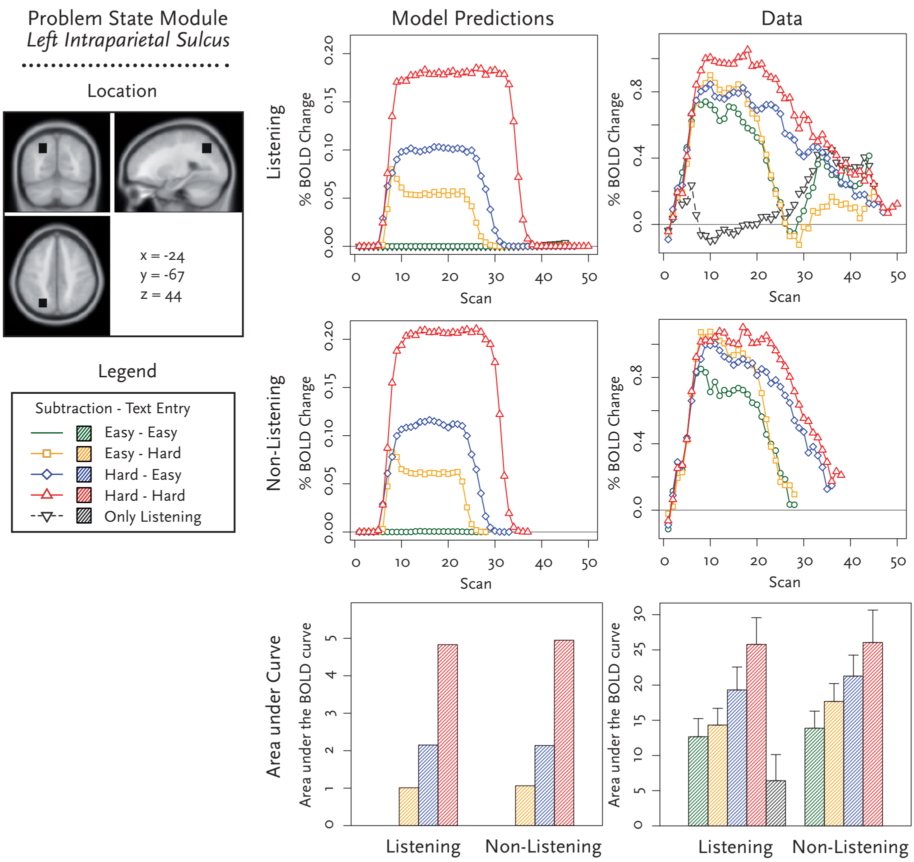


Figure 4.3 Model predictions and BOLD results for the problem state module. Please note that the black line is hidden behind the green line in the upper left graph.

It should be noted that we do not assume that modules in ACT-R cause activation in only these regions, nor that activation in these regions is only due to the associated ACT-R modules. However, these regions have been the best indicators of activation in the ACT-R modules over an extended series of studies (see also http://act-r.psy.cmu.edu/mri and Anderson, 2007).

The predictions were made using the model described above, adapted for the fMRI-suitable interface of the current experiment. While the experiment is in essence the same as Experiment 3 in Borst, Taatgen, and Van Rijn (2010), some changes were made to the interface to make it suitable for the fMRI scanner. First, in the current experiment, participants were told before each trial what the conditions of the different tasks would be to reduce noise in the fMRI measurements. This was most relevant in the difficult subtraction condition, as in the experiments in Borst, Taatgen, and Van Rijn participants only discovered during a trial that a subtraction required 'borrowing'. Second, all responses had to be made using the mouse (instead of the keyboard). Finally the interface was made more compact to reduce eye- and head movements.

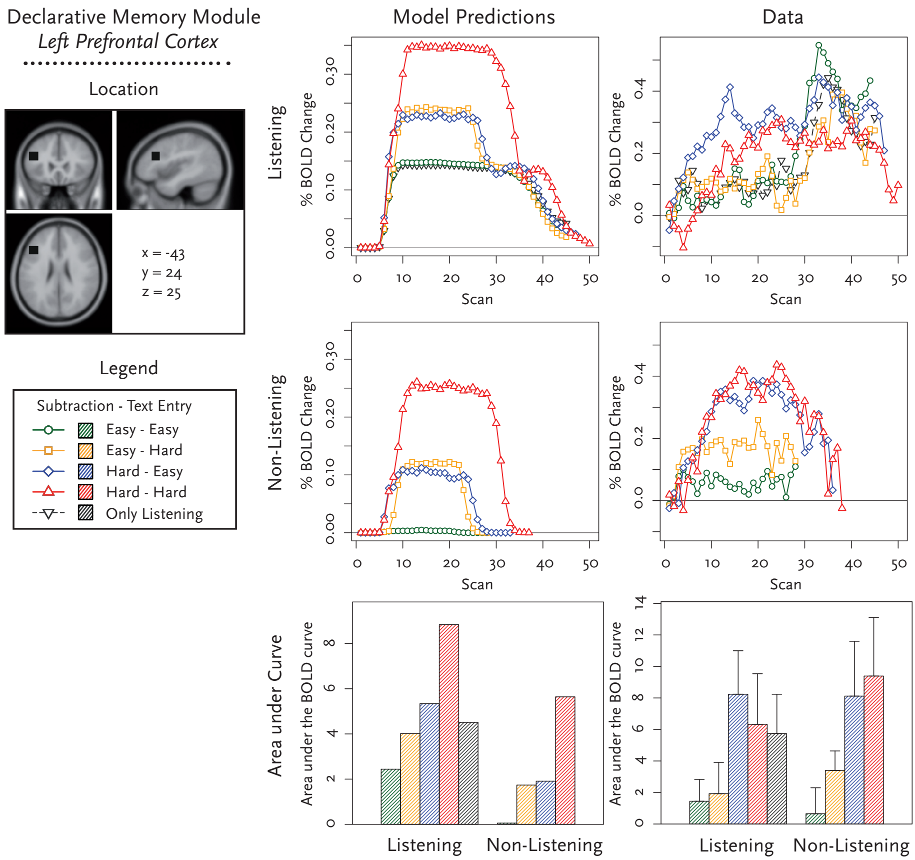


Figure 4.4 Model predictions and BOLD results for the declarative memory module. Please note that the black line is hidden behind the green line in the upper left graph.

We discuss predictions for the five most interesting modules of the current model: the problem state module, the declarative memory module, the manual module, the visual module, and the aural module (Figures 4.3–4.7; note that these predictions are based on module demand traces similar to those shown in Figure 4.2). On the left side of each figure the location and the MNI coordinates of the particular module are shown. The three graphs in the center of each figure show the model predictions; the three graphs on the right the fMRI data (which will be discussed in the 'Results' section). The four line graphs show the BOLD response over a complete trial (i.e., entering 10 digits and 10 letters, and in the case of the listening task answering the multiple-choice question). The x-axis of these graphs represents time in the form of scans (1 scan = 2 seconds); the y-axis percent BOLD change (as compared to the average of the first two scans in a trial). The two line graphs at the top show the four conditions when the listening task was present, together with the 'only listening' condition. The two line graphs in the middle show the four conditions without the listening task. Finally, the two bar graphs show the area under the curve of the BOLD graphs, indicating

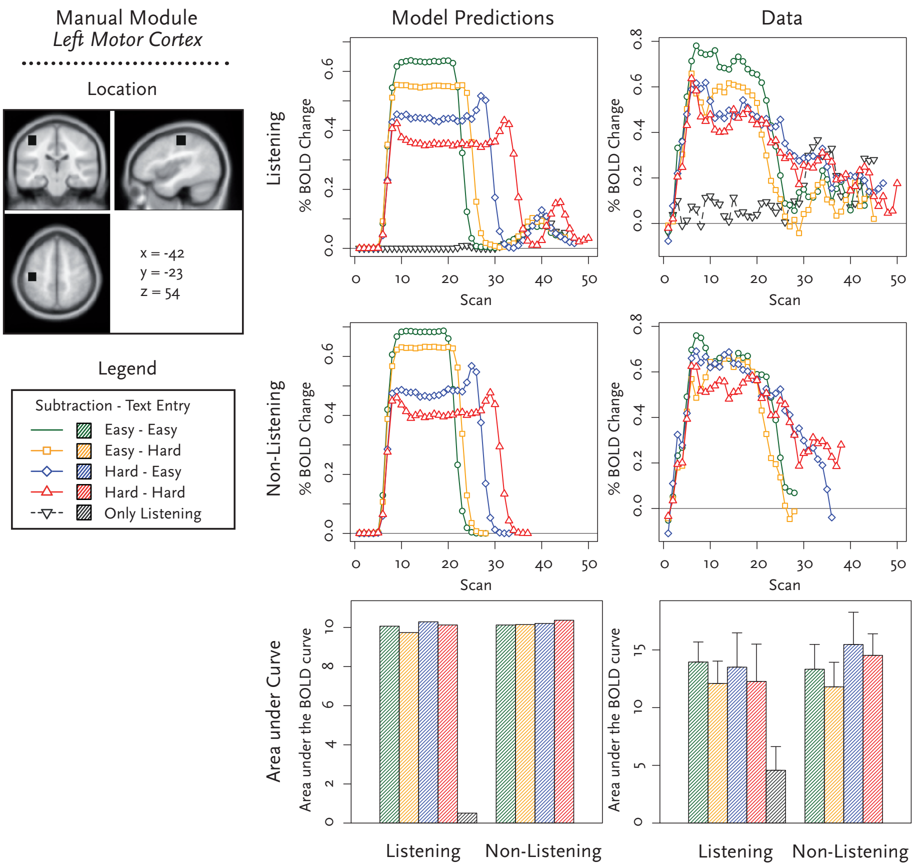


Figure 4.5 **Model predictions and BOLD results for the manual module.**

the total time a module is active and thus the total activation in a brain area during a trial (as it is sensitive to both the magnitude and the duration of the response, see Anderson, 2005; Stocco & Anderson, 2008). We will now discuss the most important predictions; lower-level predictions for each module will be discussed in the results section alongside the experimental results.

The experiment and the model were developed to investigate the problem state bottleneck. The most important prediction is therefore related to the problem state resource and its associated brain area, the intraparietal sulcus. The model claims that the problem state has to be swapped at every step in a trial in the *hard – hard* condition. In the other conditions, the problem state is either not used at all (the *easy – easy* condition), or used only for one of the tasks (*easy – hard* and *hard – easy*). Therefore, the model predicts no BOLD activity in the *easy – easy* condition, intermediate levels in the *easy – hard* and *hard – easy* conditions, and the most activity in the *hard – hard* condition (Figure 4.3; cf. Figure 4.2). In fact, as the area under the curve reflects the total time that a module is active, an over-additive interaction effect is predicted in the

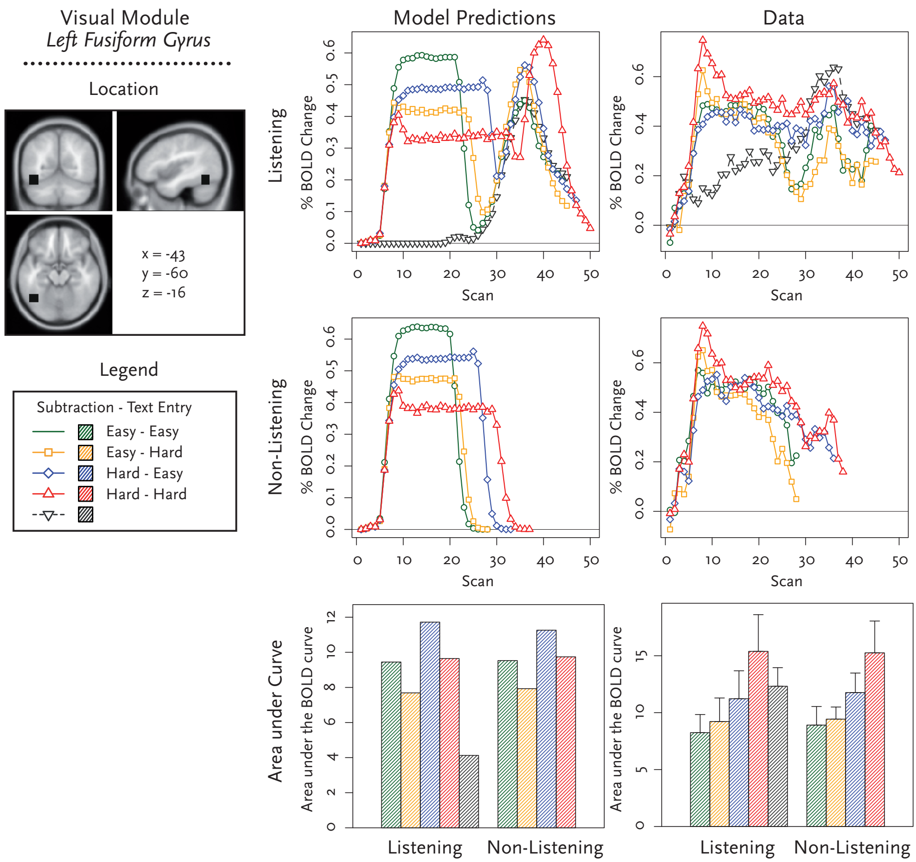


Figure 4.6 **Model predictions and BOLD results for the visual module.**

intraparietal sulcus. Because the declarative memory module is used to retrieve the old problem state on each step in a trial in the *hard – hard* condition, a similar interaction effect is predicted for the declarative memory module (Figure 4.4).

For the manual module (Figure 4.5), opposite patterns are predicted, with highest BOLD peaks occurring in the easier conditions. This may seem odd, because participants have to make the same number of responses in each condition. However, because response times are longer in the more difficult conditions, the BOLD response has more time to decay between each response (see also Figure 4.2). Therefore, the curves are lower but broader in the more difficult conditions, and higher and narrower in the easier conditions: the area under the curve is equal in all conditions. A similar pattern is predicted for the visual module (Figure 4.6). However, more visual activity is predicted for the hard subtraction condition than for the easy subtraction condition, because the model has to look multiple times at the digits to process the 'borrowings'. With respect to the aural module (Figure 4.7), the model obviously predicts no activity in the non-listening conditions, and sustained levels of activity in the listening conditions.

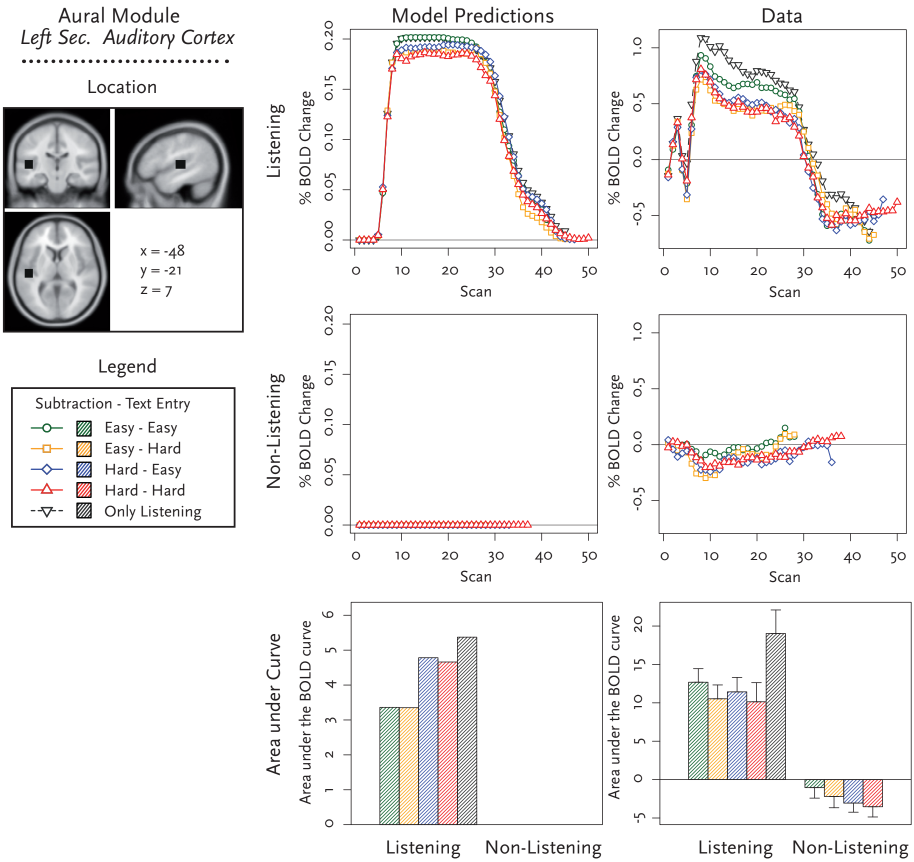


Figure 4.7 Model predictions and BOLD results for the aural module.

To summarize, the model does not predict a general increase in BOLD response with task difficulty; instead, it predicts lower but more persistent activation levels for the more difficult conditions in the visual and manual modules, and higher and more persistent activation levels for the more difficult conditions in the problem state and declarative memory modules. In the next section the fMRI experiment is described that was carried out to test those predictions.

Methods

Experimental Procedures

The design of the experiment is described in 'A Priori Predictions – The Triple Task' and Footnote 3. The participants performed the experiment in three sessions. The first session was a practice session, in which the participants were familiarized with the task, and trained for about 30 minutes. The next day the first of two fMRI sessions of about 90 minutes took place, followed by the second fMRI session a few days later (on average 3.3 days after the first session, range 1-9 days). The two fMRI sessions were identical.

Participants

Thirteen students of Carnegie Mellon University participated in the experiment. Three of them had to be excluded: one for falling asleep in the MRI scanner, one for ignoring the listening task, and one for fMRI recording problems, which leaves 10 complete datasets (3 women, average age 21.9, range 19-28, right-handed). All participants had normal or corrected-to-normal vision and normal hearing. Written informed consent as approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh was obtained before the experiment. Participants received US$ 100 compensation for performing the practice session and the two experimental sessions.

Stimuli

The stimuli for the subtraction task were generated anew for each participant. The subtraction problems in the hard version always featured six 'borrowings', and resulted in 10-digit answers. The 10 letter words for the hard version of the text-entry task were handpicked from a list of high-frequency English words (CELEX database) to ensure that similarities between words were kept at a minimum. These stimuli were also used in the easy text-entry task, except that the letters within the words were scrambled (under the constraint that a letter never appeared twice in a row). Thus, participants were presented pseudo-random sequences of letters that they had to enter one-by-one in the easy condition. By scrambling the words, we controlled for letter-based effects, while preventing the use of strategies to predict the next letter.

The audio recordings and questions for the listening task were taken from four English listening comprehension exams (university entrance-level in the Netherlands, VWO Engels 2004-2007, Cito Arnhem). The story length ranged between 26 and 72 seconds (M = 52.6, SD = 9.7). The multiple-choice questions, which participants only saw after hearing the text, had three options. These questions could be answered without making inferences, but did require attention for the complete duration of the story.

During the practice session, the experiment was presented full screen on a 17" monitor. The width of the interface measured 20 cm; the overall height 9 cm (see also Figure 4.1). Participants were sitting at a normal viewing distance, about 75 cm from the screen. The stories were presented via speakers, of which participants could control the volume using the keyboard. During the experimental sessions, the experiment was projected on a screen in the MRI scanner, allowing the participants to view the experiment via a set of mirrors attached to the head coil. The interface was operated through a normal computer mouse using the right hand. The listening task was presented via fMRI-compatible headphones, reducing scanner noise to allow the participants to hear the stories. Participants could change the volume of the stories using the mouse wheel.

Procedure

Each trial started with the presentation of a fixation cross, followed by two colored circles indicating the difficulty levels of the tasks (on the left for the subtraction task, on the right for the text-entry task; a green circle for easy, a red circle for hard, two open circles for the 'only listening' condition). If the listening task was present, a short beep sounded when the circles were displayed. The circles stayed on the screen for 5 seconds, followed by a fixation cross for 1 second. Afterwards, the subtraction and text-entry tasks appeared and, in case of the listening task, the story started. Participants always begun with the subtraction task, and then alternated between the two tasks. After completing both tasks, a feedback screen was shown for 3 seconds, indicating how many letters or digits were entered correctly. After the feedback screen and after the story was finished, the multiple-choice question was displayed. When the participants clicked on an answer, a feedback screen was shown for 4 seconds. The experiment was slow event-related, with trials separated by long breaks whose duration was sampled from a uniform distribution between 13 and 17 seconds. The onset of the circles as well as the onset of the tasks was synchronized with the beginning of a volume acquisition.

The practice session consisted of 13 single task trials, followed by a block of 9 multitask trials: all combinations of subtraction and text-entry in combination with the listening task (4 trials: *easy-easy*, *hard-easy*, *easy-hard*, and *hard-hard*), without the listening task (4 trials), and one 'only listening' trial. Both experimental sessions consisted of 5 multitask trial blocks and of one practice block at the start of a session, to re-familiarize participants with the task (this was performed during the acquisition of structural images, allowing the participants to get habituated to the environment and to adapt the listening-volume before the experimental trials). Trials were randomized within a block; stimuli were randomized over the two experimental sessions. The complete experiment (two sessions) consisted of 90 experimental trials. After each block participants could take a short break.

fMRI Procedures and Preprocessing

The fMRI data were collected with a Siemens 3T Allegra Scanner using a standard radio frequency head coil. Each functional volume existed of 34 axial slices (3.2 mm thickness, 64 × 64 matrix, 3.125 × 3.125 mm per voxel), acquired using echo-planar imaging (2000 ms TR, 30 ms TE, 79° flip angle, 200 mm field of view, 0 slice gap, with AC-PC on the 11^{th} slice from the bottom). Functional acquisition was event-related; scanning onset was synchronized with stimulus onset as described above. Anatomical images were acquired using a T1-weighted spin-echo pulse sequence at the same location as the functional images but with a finer resolution (3.2 mm thickness, 200 mm field of view, 256 × 256 matrix, 0.78125 × 0.78125 mm in-plane resolution).

Table 4.2 ANOVA results of the text-entry task.

	Response Times			Accuracy		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
Listening	10.69	.010	.54	10.37	.010	.54
Subtraction	32.43	< .001	.78	8.72	.016	.49
Text-Entry	5.67	.041	.39	32.17	< .001	.78
Listening × Subtraction	< 1	–	–	1.60	.24	.15
Listening × Text-Entry	< 1	–	–	< 1	–	–
Subtraction × Text-Entry	12.08	.007	.57	10.35	.01	.53
Listening × Sub. × Text-Entry	< 1	–	–	< 1	–	–

Subtraction = Subtraction Difficulty, Text-Entry = Text-Entry Difficulty.

Table 4.3 ANOVA results of the subtraction task.

	Response Times			Accuracy		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
Listening	1.58	.24	.15	< 1	–	–
Subtraction	83.82	< .001	.90	80.96	< .001	.90
Text-Entry	5.40	.045	.38	4.34	.067	.33
Listening × Subtraction	< 1	–	–	1.44	.26	.14
Listening × Text-Entry	< 1	–	–	2.81	.13	.24
Subtraction × Text-Entry	2.70	.13	.23	14.05	.005	.61
Listening × Sub. × Text-Entry	1.47	.26	.14	< 1	–	–

Subtraction = Subtraction Difficulty, Text-Entry = Text-Entry Difficulty.

The data were analyzed using SPM5 (Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/). This included realigning the functional images, coregistering them with the structural images, normalizing the images to the MNI (Montreal Neurological Institute) ICBM 152 template, and smoothing them with an 8 mm FWHM Gaussian kernel. The MarsBaR toolbox (Brett, Anton, Valabregue, & Poline, 2002) was used to extract the time course information in predefined regions.

Results

We will first discuss the behavioral results, followed by the fMRI region-of-interest results. An exploratory fMRI analysis was also performed, which confirmed the existence of peaks of activations in the standard ACT-R regions-of-interest (see Appendix 1 and Tables 4.6, 4.7, and 4.8 for more details). All reported *F*- and *p*-values are from repeated measure analyses of variance (ANOVAs), all error bars depict standard errors, effects were judged significant when a .05 significance level was reached, and accuracy data were transformed using an arcsine transformation before performing ANOVAs.

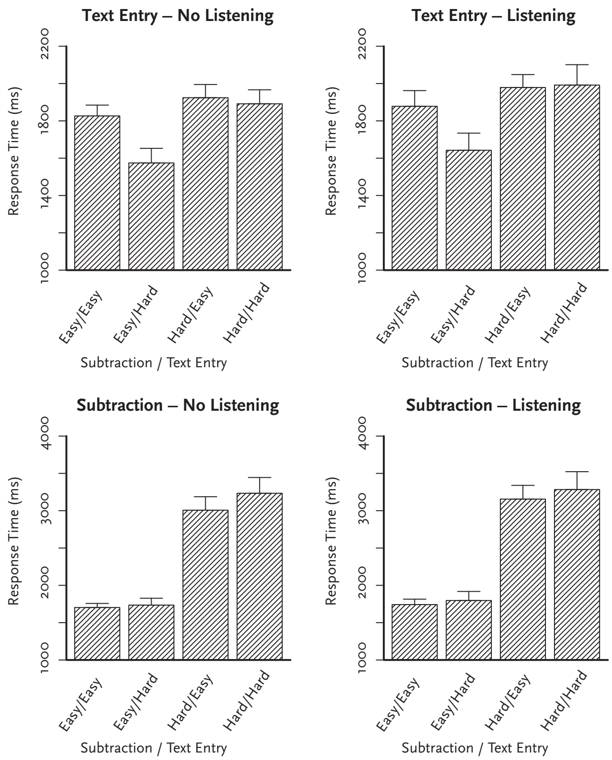


Figure 4.8 Response times on the subtraction and text-entry tasks. Error bars represent standard errors.

Behavioral Results

Outliers in response times were eliminated by means of a two-step procedure. First, response times faster than 250 ms and slower than 10,000 ms were removed. Then, data exceeding 3 standard deviations from the mean per condition per participant were excluded. Overall, 2.4% of the data was discarded. Table 4.2 (text-entry) and Table 4.3 (subtraction) summarize the results.

Results

Figure 4.8, upper panels, shows the response times on the text-entry task, on the left without and on the right in combination with the listening task. A response time on the text-entry task was defined as the time between entering a digit in the subtraction task and entering a letter in the text-entry task. The first responses of each trial were removed (per task), as they might contain 'start-up' effects. An ANOVA showed that all three main effects were significant (see Table 4.2), indicating that response times decreased with Text-Entry Difficulty, but increased with Subtraction Difficulty and Listening. The interaction between Subtraction Difficulty and Text-Entry Difficulty also

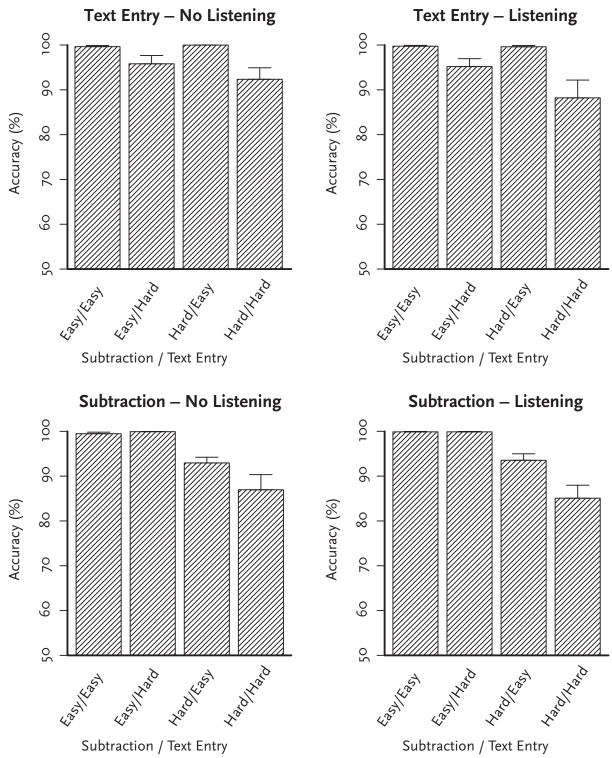


Figure 4.9 Accuracy on the subtraction and text-entry tasks. Error bars represent standard errors.

reached significance, which is due to the increased response times in the *hard-hard* condition, as was predicted. The three-way interaction did not reach significance.

The lower panels of Figure 4.8 show the response times on the subtraction task. This is the time between clicking a button in the text-entry task and entering a digit in the subtraction task. Again, the first response of each trial was removed. Only the main effects of Subtraction Difficulty and Text-Entry Difficulty reached significance (see Table 4.3), showing an increase in response times for both effects. The interaction of Subtraction Difficulty and Text-Entry Difficulty did not reach significance, nor did any effects involving the listening task.

Figure 4.9 shows the accuracy on the subtraction and text-entry tasks (the ANOVA results are listed in Table 4.2 and 4.3). The two top panels show the accuracy on the text-entry task. All three main effects reached significance, all three indicating a decrease in accuracy. As predicted, when both the subtraction and the text-entry task were hard, accuracy decreased even more, which is shown by the significant interaction between Subtraction Difficulty and Text-Entry Difficulty. The other effects did not reach significance. The lower panels of Figure 4.9 show the accuracy on the subtraction task. The main effect of Subtraction Difficulty was significant, as was the interaction between Subtraction Difficulty and Text-Entry Difficulty. The other tests did not reach significance.

Figure 4.10 shows the accuracy on the listening task. One of the stories was removed because participants' accuracy was at chance level. Only the main effect of Subtraction Difficulty reached significance ($F(1,9) = 18.09$, $p = .002$, $\eta_p^2 = .67$), caused by a decrease in accuracy when subtraction was hard. The other effects were not significant (Text-Entry Difficulty: $F < 1$; Subtraction Difficulty × Text-Entry Difficulty: $F(1,9) = 1.29$, $p = .28$, $\eta_p^2 = .13$).

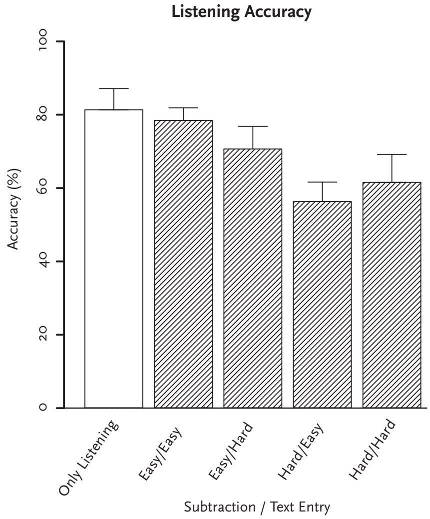


Figure 4.10 Accuracy on the listening task. Error bars represent standard errors.

Discussion

The results were as expected: the interaction effect between Subtraction Difficulty and Text-Entry Difficulty was significant for the response times of the text-entry task and for the accuracy scores of both tasks. Thus, when a problem state was required for both tasks (the *hard-hard* condition), response times increased and accuracy decreased, as was predicted by the model. The fact that this interaction did not reach significance ($F(1,9) = 2.7$, $p = .13$, $\eta_p^2 = .23$) for the response times of the subtraction task is probably due to the lower number of participants than in previous experiments, in which the effect was always significant (Borst, Taatgen, & Van Rijn, 2010). Furthermore, compared to previous experiments, response times were slightly higher. This difference is probably due to performing the experiment in the scanner and using the mouse.

The pattern of response times of the text-entry task was slightly different than in the previous experiment (Experiment 3 of Borst, Taatgen, & Van Rijn, 2010): response times were lower in the hard version of the text-entry task than in the easy version. The explanation is that participants have to do two different actions to determine the next letter to type in the text-entry task: in the easy version they have to look at the letter that they need to type, and in the hard version they have to mentally determine the next letter to type given the word and position. In earlier experiments these two actions happened to take approximately the same amount of time, but in the current experiment the action for the easy version of the task turned out to be slower, probably due to the slightly different interface that we used in this experiment. We have observed similar effects before, for instance in Experiment 2 of Borst, Taatgen, and Van Rijn. To ensure that the fMRI experiment is comparable to our earlier studies, we ran the same experiment outside the scanner. This yielded similar results as reported previously (Borst, Taatgen, & Van Rijn, 2010), including the decrease of response times for the hard text-entry task, indicating that the observed differences are not due to the minor changes in the task interface. That is, the fact that the interaction effect did not reach significance is probably caused by the low number of participants, and the slightly different pattern of results by performing the task lying in the scanner and operating the mouse in this setup. This suggests that the experiment still taps the same underlying cognitive constructs and that it is therefore comparable to our previous studies. For details see Appendix 2, Figures 4.11 and 4.12, and Tables 4.9 and 4.10.

The model did not predict the effect of the listening task on the text-entry task, neither the small increase in response time nor the small decrease in accuracy. As this is not the focus of the current paper, we did not pursue this issue further. The effect of the subtraction task on the listening task accuracy, Figure 4.10, is explained by the model: When the subtraction task is hard, there is a high demand for declarative memory, causing the model to not process all words of the listening task (for which declarative memory is also required). This could then lead to more mistakes in answering the questions (see Borst, Taatgen, & Van Rijn, 2010, for a more extensive discussion of this issue).

fMRI Results: Regions-of-Interest

To analyze the effects in the predefined regions, we first transformed the Talairach-Tournoux coordinates used in previous ACT-R/fMRI papers (e.g., Anderson, 2007) to the MNI coordinates reported in Table 4.1 using a non-linear mapping (Lacadie, Fulbright, Rajeevan, Constable, & Papademetris, 2008). The smoothed functional images were proportionally and grand mean scaled (with a grand mean of 100) using SPM. The BOLD response was then calculated as percent signal change as compared to the first two scans of a trial. Trials belonging to the same participant, brain area, and condition were averaged together. Because the area under the curve reflects the total activity of a brain area (see Anderson, 2005; Stocco & Anderson, 2008), we entered this value into an ANOVA. We only took the area between the start of a trial and the behavioral feedback screens into account, because the tails of the BOLD curves contain the multiple-choice questions. These could obscure the results and were not included in the ACT-R model (except for the passive act of reading the words on the screen). Table 4.4 contains the results; Table 4.5 shows which scans were taken into account for the different conditions.

Results

The most important prediction of the model was an over-additive interaction effect in the intraparietal sulcus, reflecting the problem state bottleneck. Figure 4.3 shows the results in the intraparietal sulcus: most activation is indeed observed for the *hard- hard* condition. The ANOVA of the area under the curve shows that the interaction between Subtraction Difficulty and Text-Entry Difficulty is significant in combination with the listening task, but not without it (see Table 4.4). Furthermore, the main effects of Subtraction Difficulty and Text-Entry Difficulty are significant with and without the Listening task. The model prediction that there is no activation for the *easy-easy* condition did not come true, but the prediction that the problem state resource is not

Table 4.4 **ANOVA results of the area under the curve of the regions-of-interest.**

Problem State Module – Intraparietal Sulcus						
	With Listening			*Without Listening*		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
Subtraction	60.20	< .001	.87	20.89	.001	.70
Text-Entry	6.49	.031	.42	10.39	.010	.54
Subtraction × Text-Entry	5.15	.049	.36	< 1	–	–
Declarative Memory Module – Prefrontal Cortex						
	With Listening			*Without Listening*		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
Subtraction	11.73	.008	.57	7.53	.023	.46
Text-Entry	< 1	–	–	9.81	.012	.52
Subtraction × Text-Entry	2.03	.19	.18	< 1	–	–
Manual Module – Motor Cortex						
	With Listening			*Without Listening*		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
Subtraction	< 1	–	–	3.51	.09	.28
Text-Entry	1.85	.207	.17	< 1	–	–
Subtraction × Text-Entry	< 1	–	–	< 1	-–	–
Visual Module – Fusiform Gyrus						
	With Listening			*Without Listening*		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
Subtraction	11.52	.008	.56	11.12	.009	.55
Text-Entry	3.00	.117	.25	4.06	.075	.31
Subtraction × Text-Entry	2.42	.154	.21	1.35	.276	.13
Aural Module – Secondary Auditory Cortex						
	With Listening			*Without Listening*		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
Subtraction	< 1	–	–	3.32	.102	.27
Text-Entry	2.00	.191	.18	1.39	.269	.13
Subtraction × Text-Entry	< 1	–	–	< 1	–	–

used for the listening task – except for answering the multiple-choice question – is reflected by the data.

Figure 4.4 shows the results of the prefrontal cortex, associated with the retrieval module. For this region the model also predicted an over-additive interaction effect of Subtraction Difficulty and Text-Entry Difficulty, which was not found in the data. The model also predicted main effects of both Subtraction Difficulty and Text-Entry Difficulty and these effects were indeed found (Text-Entry Difficulty was only significant without the listening task, Subtraction Difficulty both with and without listening).

Table 4.5 Number of scans that was taken into account for the analyses of the area under the curve, per condition.

Subtraction	Text-Entry	Listening	Scans
Easy	Easy	No	23
Easy	Hard	No	23
Hard	Easy	No	31
Hard	Hard	No	33
Easy	Easy	Yes	23
Easy	Hard	Yes	24
Hard	Easy	Yes	32
Hard	Hard	Yes	34

In contrast to the problem state and declarative memory modules, we expected a higher BOLD response peak for the easier conditions in the manual and visual areas. Indeed, in the motor cortex – associated with the manual module – the BOLD curve reached its highest activation levels in the *easy-easy* condition (Figure 4.5). The more difficult the condition, the lower and broader the activation curves. The model predicted no effects on the total activity; this was confirmed by the data.

At first sight, the match between model and empirical data for the fusiform gyrus (Figure 4.6), associated with the visual module, seems less convincing. However, a more careful analysis shows that the same patterns are observed in both model and data. The model predicted an effect of Subtraction Difficulty on activation in the fusiform gyrus, as the digits have to be visually attended to multiple times in the hard condition to solve the 'borrowings'. This is confirmed by the ANOVA that compared the area under the curve between the easy and difficult conditions. While the model also predicted a small decrease of visual activation in the hard text-entry conditions (because in the hard condition the word only had to be read at the first step of a trial, while in the easy condition a letter had to be processed at each step of a trial), this was not found in the data. Finally, the model predicted a peak of activation around scan 40 caused by reading the multiple-choice questions; this was reflected in the data.

Figure 4.7 illustrates the results for the auditory cortex. As expected, when the listening task was not present, the BOLD response was absent. When the listening task was present, on the other hand, a clear BOLD response was found. The model predicted this effect, and additionally predicted a small effect of condition. The cause of this effect is that to process each word that the model hears, it has to retrieve multiple facts from declarative memory. In the more difficult subtraction and text-entry conditions, these tasks also make heavy demands on declarative memory. When declarative memory is busy, the model can sometimes not process a word right away, which results in missing some words in the auditory stream. Thus, the more difficult the subtraction and text-entry conditions, the higher the demands on declarative memory, the more words are missed, and the lower the predicted BOLD response for the secondary auditory cortex (as this region reflects processing auditory information, not passive listening). A similar effect seems to be present in the data, but did not reach significance.

Discussion

The atypical prediction that the more difficult conditions would show lower but broader activation curves in the visual and manual regions, and higher and broader curves in the problem state and declarative memory conditions, was confirmed by the data. Furthermore, the over-additive interaction effect in the problem state region was present in the fMRI data (in combination with the listening task), supporting the theory that the problem state bottleneck is localized in the intraparietal sulcus. This interaction effect was not found in the declarative memory region (see the General Discussion for an extensive discussion of this issue). In the aural region the predictions were confirmed in general: a BOLD response in the listening conditions, mediated by the conditions of the subtraction and text-entry tasks. However, these effects did not reach significance.

General Discussion

The current study was performed to investigate the neural correlates of problem states and the problem state bottleneck, and to validate our theory using neuroimaging data. First, we generated *a priori* fMRI predictions for five brain areas using our model, which were subsequently tested in an experiment. This resulted in two main predictions: (1) an over-additive interaction effect in the problem state region (the intraparietal sulcus) and in the declarative memory region (a part of the prefrontal cortex), and (2) lower and broader BOLD curves for the more difficult conditions in the manual and visual regions, and higher and broader BOLD curves for the more difficult conditions in the problem state and declarative memory regions. The first prediction came true for the problem state region, but not for the declarative memory region, while the counter-intuitive second prediction was confirmed by the experiment.

In general, the model's fMRI predictions for this complex task were accurate. The paper focuses mainly on the overall BOLD response in the regions (area under the curve). The figures also report the time course of the BOLD response over a trial together with the corresponding model predictions. Here the fit between the scan-by-scan data points and the model is more modest, which can be explained by the fact that we made *a priori* predictions, and did not try to fit the curves post-hoc. A number of factors might be called into question. First, ACT-R uses only a simple gamma function, identical for every module, to predict the BOLD response in each region. However, the biological hemodynamic response function is more complex than that, and varies in different parts of the brain (e.g., Handwerker, Ollinger, & D'Esposito, 2004). Choosing different functions and fitting their parameters for each region separately would probably result in a better model-data match. Second, due to the duration of our experimental paradigm, only a relatively small number of observations for each condition was available for each participant. This small number of observations might not be able to cancel the scan-to-scan variations of noise in the MRI signal, thus making the true shape of the observed BOLD curves difficult to estimate. Third, the model might be underestimating some trial-by-trial variability in the subjects' responses. In particularly long trials, the BOLD response in a region might cumulate over the interval between trials and carry over to the scans chosen as a baseline for the successive trial. The fact that certain BOLD curves (especially in Figures 4.4 and 4.5) do not return to baseline suggests that this kind of contamination was indeed occurring, possibly corrupting the true shape of the BOLD curves.

It should be noted that, while all these factors can affect the shape of the BOLD response, none of them should significantly impact our predictions on the relative magnitudes of the areas under the curve. Therefore we choose to focus on the predictive power of the model and its principal predictions. In combination with the behavioral evidence that we gathered before (Borst, Taatgen, & Van Rijn, 2010), the observed global effects on the BOLD response suggest that the hypothesized existence of a problem state bottleneck can explain the interference effects in the data.

The most important prediction of the model was an over-additive interaction effect in the problem state region. While this effect was indeed present in the data in combination with the listening task, it did not reach significance in the trials without the listening task. The main reason for this is that while the model predicted no activity in the problem state region for the *easy-easy* condition, the experimental data does show increased activity in this condition. One possible explanation is that the observed activity was caused by visually processing the stimuli, as the same parietal region is known to be involved in visual-spatial processing (e.g., Culham & Kanwisher, 2001). Not only would this lead to an effect in the *easy-easy* condition, but also obscure the effects in the other conditions. In combination with non-linear properties of the BOLD response (e.g., Bandettini & Ungerleider, 2001; Dale & Buckner, 1997; Vazquez & Noll, 1998), this could explain why we did not observe the interaction effect here, especially taken into account the relatively low number of participants. Another possibility is that participants use their problem state resources in the *easy-easy* condition to represent information, even if this is not required by the task. This would lead to neural activity in the easy-easy condition, again canceling the interaction effect. The additional load of the listening task could have prevented the use of problem state resource (see, for similar effects, Taatgen et al., 2009), which could explain why we did find the interaction effect in the context of the listening task. However, as the model has successfully accounted for data of three experiments (Borst, Taatgen, & Van Rijn, 2010), we do believe that the basic mechanisms of the model are sound, and decided against post-hoc changes to the model.

The prefrontal region corresponding to the declarative memory module exhibits the predicted main effects of subtraction difficulty and text-entry difficulty (except for Text-Entry Difficulty when the listening task was present). This supports the hypothesis that this predefined region indeed represents an area involved in the processing of declarative memory elements (such as subtraction facts). However, we did not find the predicted interaction effect. The interaction effect was supposed to be caused by encoding of problem states (on top of retrieving subtraction facts). Even though the predefined area is known to be active when intentionally encoding facts and even when unintentionally encoding facts (e.g., Buckner, Kelley, & Petersen, 1999), the experiment did not provide evidence that it is actually used to encode suspended problem states. Therefore, either this region's contribution to the processing problem states was too weak to impact the BOLD signal, or the retrieval of suspended problem states is controlled by a different region.

With respect to the first option (i.e., the contribution to the signal being too weak) one must note that the predictions made by our model were based on the assumption that both retrieving a previous problem state from declarative memory and swapping it into the problem state module require some measurable cost in terms of time. When the model was fit to the behavioral data of Borst, Taatgen, and Van Rijn (2010), these two costs had to be estimated together, with no possibility of disentangling them. However, it is conceivable that the retrieval time for a problem state is very short, and that most of the time is due to the swapping process. Under such circumstances, the model would still predict the over-additive effects of task difficulty for the problem state region, but not for the retrieval region. In fact, there are at least two reasons why the retrieval time for problem states should be very short. The first reason is recency: the problem state that needs to be retrieved has been swapped out of its module only a few seconds before, and it is probably still active in memory. Second, the retrieval of appropriate problem states can be easily cued by task-relevant, on-screen information. In both cases, there is no reason to expect a significant effect of problem state retrievals on the prefrontal region. In fact, the pattern of data in Figure 4.5 (lower half) suggests that main factor affecting the response of the retrieval region is the difficulty of the subtraction task. Thus, although the interaction effect in the PFC was an a priori prediction of our model, it was not an inevitable prediction and its lack does not undermine the plausibility of our framework.

On the basis of previous ACT-R/fMRI research, the model predicted that the problem state resource – and thus the effect of the bottleneck – is located in the intraparietal sulcus. This notion is supported by the current results: the predicted interaction effect caused by the problem state bottleneck was found in this region. This region is part of the fronto-parietal network that is consistently found in neuro-imaging studies of working memory. While the intraparietal sulcus is mostly implicated in spatial working memory and spatial attention tasks, it is also known to be responsible for object and verbal working memory (among other regions, e.g., LaBar, Gitelman, Parrish, & Mesulam, 1999; Wager & Smith, 2003). In our study, the problem states did not contain spatial information, and therefore confirms a more general role of the intraparietal sulcus for working memory.

In the hard subtraction task the problem state resource contained numerical information, that is, information whether a 'borrowing' is in process. It is not surprising that this leads to increased activation in the intraparietal sulcus, as the horizontal part of the intraparietal sulcus is one of the three circuits for numerical processing as identified by Dehaene, Piazza, Pinel, and Cohen (2003). In the hard text-entry task, the problem state is used to maintain verbal information. Brodmann Area 40, a region bordering on the intraparietal sulcus, is known to be involved in verbal working memory, specifically in maintaining verbal working memory (e.g., Smith, Jonides, Marshuetz, & Koeppe, 1998), and it is thus also not surprising that this region is involved in maintaining the problem state for the text-entry task. While slightly different regions are implicated for storage of different kinds of information, this study suggests that maintaining more than one problem state of any kind at a time results in significant interference.

The current results also seem to suggest that the problem state is modality-specific. The subtraction and text-entry task elicit activation in the intraparietal sulcus even when they are easy (while a problem state is not required), and interfere with each other in the hard – hard condition. The listening task, on the other hand, hardly causes activation in the intraparietal sulcus (as shown by the 'only listening' condition), nor does it cause multitasking interference. As the listening task is the only non-visual task, this could imply that the intraparietal sulcus is only involved in maintaining visual problem states (cf. Anderson, Qin, Jung, & Carter, 2007).

In the current mapping scheme of ACT-R processes to brain regions, the problem state predicts activation as a function of problem state transformations, but not in reaction to storing problem states. This may seem odd, as storing problem states should also have metabolic costs. In practice, however, the two processes of storing and manipulating are difficult to separate, (as storing always follows a problem state manipulation) and previous research (see e.g., Anderson, 2007) has led to estimated costs for transformations only, assuming that representations persist at no additional metabolic cost. Therefore, we decided to keep our model as parsimonious as possible and not to introduce ad-hoc estimates of the storing costs for problem states.

Our model is based on threaded cognition (Salvucci & Taatgen, 2008, 2011; Salvucci, Taatgen, et al., 2009), a theory of multitasking that assumes multiple central and peripheral processing bottlenecks. This in contrast to for instance the EPIC theory (Meyer & Kieras, 1997a), which assumes only peripheral bottlenecks, and the central bottleneck theory (e.g., Pashler, 1994), which assumes only a single central bottleneck. While brain evidence for a central bottleneck in the frontal lobes has been reported before (e.g., Dux, Ivanoff, Asplund, & Marois, 2006), the current fMRI results give evidence for an additional central bottleneck located in the intraparietal sulcus, corroborating multiple-bottleneck theories.

In conclusion, this study lends additional support to the notion of the problem state bottleneck. This bottleneck can cause considerable interference not only in concurrent multitasking – as most bottlenecks – but also in sequential multitasking: When multiple alternating tasks need to store intermediate results, the problem state bottleneck will cause significant interference. Take for instance the prototypical example of taking a phone call while working on a paper: if you had a sentence in mind before taking the call, you will almost certainly have forgotten about it after the call.

Appendix 1: Exploratory fMRI Analysis

An exploratory analysis was performed to identify regions that responded significantly to our experimental manipulations. The results of this analysis are reported here. All analyses were performed using the general linear model implemented in SPM5. The colored circles indicating task condition (see the Method – Procedure section, first paragraph), the 'real' trial, and the feedback presentations were modeled for each condition separately. Realignment parameters were included as covariates and a high-pass filter with a 128 sec cutoff was applied. For each voxel, the hemodynamic response function (HRF) and its time and dispersion derivatives were fitted. Contrast images for each condition were made for the individual participants, and entered into second level random-effect group analyses. The statistical results were thresholded using a false-discovery-rate (FDR) correction for multiple comparisons of 0.05 and more than 40 contiguous voxels. The results are summarized in Table 4.6 (Listening > Non-Listening), Table 4.7 (Hard Text-Entry > Easy Text-Entry), and Table 4.8 (Hard Subtraction > Easy Subtraction). We also tested the interaction of Subtraction Difficulty and Text-Entry Difficulty, however, no brain areas survived the significance test.

Results

Table 4.6 lists the areas with greater activation when all three tasks had to be performed as compared to when only the subtraction and text-entry tasks had to be performed, reflecting activity related to the listening task. As expected, activation was found in bilateral temporal areas and in the left inferior frontal gyrus (cf. Just, Keller, & Cynkar, 2008). According to for instance Friederici (Friederici, 2002), the temporal regions perform identification processes, while the left frontal area integrates the words and sentences into a coherent whole. The temporal region overlaps with the area associated with ACT-R's aural module (see also Figure 4.9).

In Table 4.7 areas are listed where more activation was found when the text-entry task was hard as compared to when it was easy, thus when a problem state was required as compared to when it was not. The first area exists of large parts of the bilateral superior and inferior parietal lobules, including the intraparietal sulcus. This region includes the predefined problem state region, and was therefore expected to show an effect. The region is associated with attention and the integration of information (e.g., Culham & Kanwisher, 2001), and it is therefore not surprising that it is stronger activated by the increase in task-coordination that is necessary to perform the hard text-entry task. A second large activated network was found in the medial frontal cortex: the left Supplementary Motor Area (SMA), the left superior medial gyrus, extending into the left and right precentral gyri (the region extends into the left inferior frontal gyrus, which we will discuss below). This network is presumably active in response to the increase of cognitive control that is necessary for the hard text-entry task in combination with the subtraction task (e.g., Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). Furthermore, the middle and inferior frontal gyri were active bilaterally. These regions are very close to ACT-R's declarative memory

Table 4.6 Exploratory analysis results. Areas with greater activation for Subtraction, Text-Entry, and Listening than Subtraction and Text-Entry alone ($p < .05$, FDR corrected, >40 contiguous voxels).

Gray matter of peak activation	Size in voxels	$t(27)$	MNI coordinates
R Superior Temporal Gyrus	1575	19.03	63, -12, 0
L Superior/Middle Temporal Gyrus	2055	15.5	-57, -18, 3
L Inferior Frontal Gyrus	78	6.24	-54, 24, 12

Voxels are 3 × 3 × 3 mm. MNI = Montreal Neurological Institute.

Table 4.7 Exploratory analysis results. Areas with greater activation for Hard Text-Entry than Easy Text-Entry ($p < .05$, FDR corrected, >40 contiguous voxels).

Gray matter of peak activation	Size in voxels	$t(27)$	MNI coordinates
L SMA / L Superior Medial Gyrus / L Inferior Frontal Gyrus	1287	9.16	-3, 6, 60
R Middle Frontal Gyrus	318	8.18	42, 36, 33
L & R Superior / Inferior Parietal Lobules	1115	7.49	-9, -69, 51
L Middle Frontal Gyrus	233	5.94	-42, 45, 21
R Insula Lobe	123	5.28	30, 27, 0

Voxels are 3 × 3 × 3 mm. MNI = Montreal Neurological Institute. SMA = Supplementary Motor Area.

Table 4.8 Exploratory analysis results. Areas with greater activation for Hard Subtraction than Easy Subtraction ($p < .05$, FDR corrected, >40 contiguous voxels).

Gray matter of peak activation	Size in voxels	$t(27)$	MNI coordinates
R Superior Frontal Gyrus	113	7.61	27, 12, 57
R Supra Marginal Gyrus / R Inferior Parietal Lobule	485	7.58	48, -39, 45
R Middle Frontal Gyrus	215	6.69	45, 42, 24
L Middle / Inferior Frontal Gyrus	301	6.53	-48, 36, 24
L Inferior Parietal Lobule	317	6.48	-36, -48, 39
L SMA	90	5.50	3, 21, 45

Voxels are 3 × 3 × 3 mm. MNI = Montreal Neurological Institute. SMA = Supplementary Motor Area.

area, reflecting an increase in memory retrievals necessary for interpreting words and spelling information.

A very similar set of regions was found when we compared the hard subtraction task to the easy condition: a bilateral parietal network, a control network around the medial frontal cortex, and a memory network in the middle and inferior frontal gyri (Table 4.8). An increase of cognitive control and information processing requirements is even clearer for the subtraction task than for the text-entry task: more difficult subtraction facts have to be retrieved and 'borrowings' have to be processed.

Appendix 2: Behavioral Results outside the Scanner

Here we report the results of the experiment that we ran outside the fMRI scanner. This experiment was performed to test whether differences between the behavioral results of the current fMRI experiment and the previous experiment (Experiment 3, Borst, Taatgen, & Van Rijn, 2010) are due to performing the experiment in the scanner and the low number of participants, or to the slightly different interface.

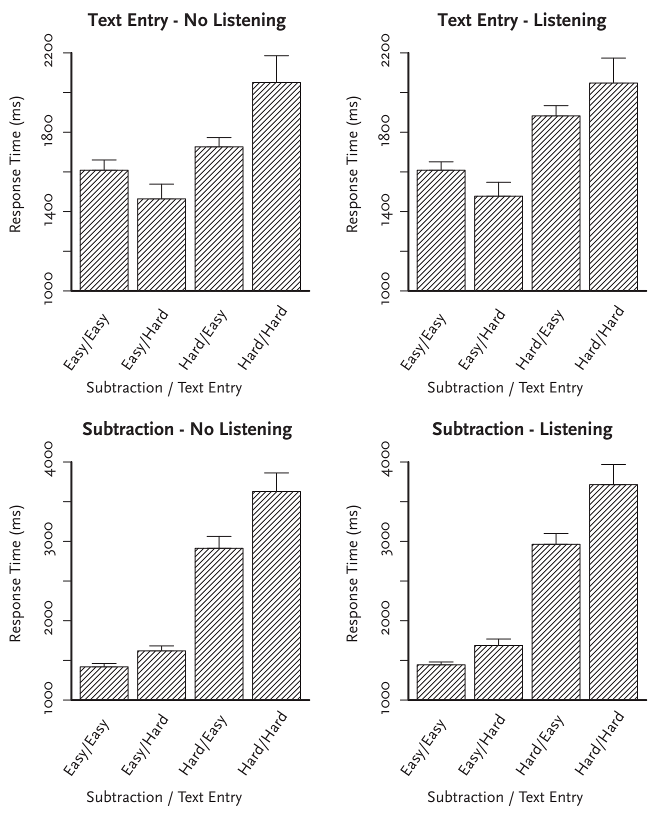


Figure 4.11 Response times outside the scanner. Error bars represent standard errors.

Table 4.9 **ANOVA results of the Text-Entry Task outside the scanner.**

	Response Times			Accuracy		
Source	$F(1,19)$	p	η_p^2	$F(1,19)$	p	η_p^2
Listening	1.50	.236	.07	< 1	–	–
Subtraction	69.49	< .001	.79	15.62	< .001	.45
Text-Entry	< 1	–	–	38.63	< .001	.67
Listening × Subtraction	1.62	.219	.08	2.00	.173	.10
Listening × Text-Entry	1.37	.256	.07	< 1	–	–
Subtraction × Text-Entry	19.85	< .001	.51	16.03	< .001	.46
Listening × Sub. × Text-Entry	2.57	.126	.12	3.06	.096	.14

Subtraction = Subtraction Difficulty, Text-Entry = Text-Entry Difficulty.

Table 4.10 **ANOVA results of the subtraction task.**

	Response Times			Accuracy		
Source	$F(1,19)$	p	η_p^2	$F(1,19)$	p	η_p^2
Listening	< 1	–	–	< 1	–	–
Subtraction	139.23	< .001	.88	52.86	< .001	.74
Text-Entry	31.99	< .001	.63	6.10	.023	.24
Listening × Subtraction	< 1	–	–	< 1	–	–
Listening × Text-Entry	< 1	–	–	1.93	.181	.09
Subtraction × Text-Entry	22.09	< .001	.54	4.19	.055	.18
Listening × Sub. × Text-Entry	< 1	–	–	1.26	.276	.06

Subtraction = Subtraction Difficulty, Text-Entry = Text-Entry Difficulty.

Results

Participants

Twenty students of Carnegie Mellon University participated in the experiment (11 women, average age 20.6, range 18–23). All participants had normal or corrected-to-normal vision and normal hearing. Informed consent as approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh was obtained before the experiment. Participants received US$ 10 for performing the experiment.

Results

Outliers in reaction times were eliminated by means of a two step procedure. First, response times faster than 250 ms and slower than 10,000 ms were removed. Then, data exceeding 3 standard deviations from the mean per condition per participant were excluded. Overall, 2.0% of the data was discarded. Figure 4.11 (response times) and Figure 4.12 (accuracy) show the results; Table 4.9 (text-entry) and Table 4.10 (subtraction) list the results of the analyses. All reported *F*- and *p*-values are from repeated measure analyses of variance (ANOVAs), all error bars depict standard errors, effects were

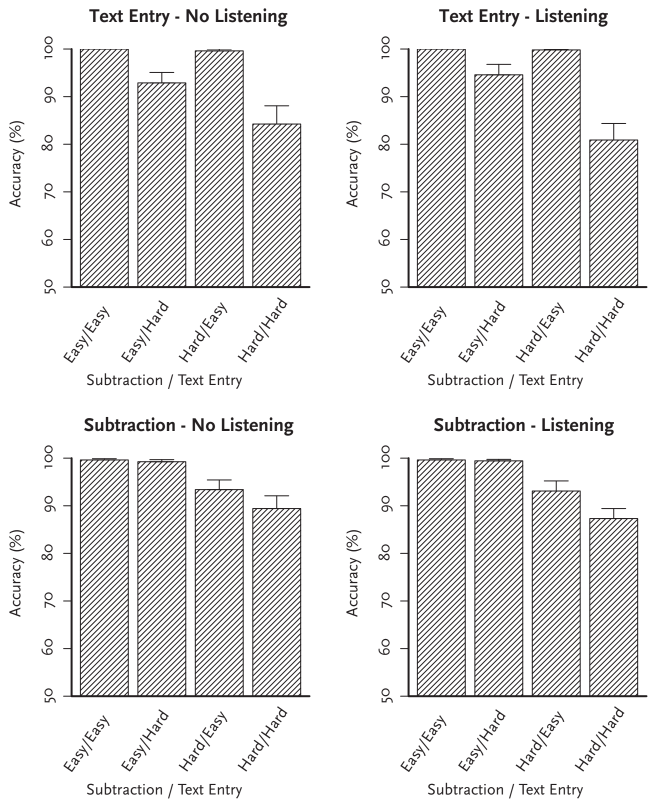


Figure 4.12 **Response times outside the scanner. Error bars represent standard errors.**

judged significant when a .05 significance level was reached, and accuracy data were transformed using an arcsine transformation before performing ANOVAs.

Figure 4.11, upper panels, shows the response times on the text-entry task, on the left without and on the right in combination with the listening task. A response time on the text-entry task was defined as the time between entering a digit in the subtraction task and entering a letter in the text-entry task. The first responses of each trial were removed (per task), as they might contain 'start-up' effects. An ANOVA showed that the main effect of Subtraction Difficulty was significant (see Table 4.9), indicating that response times increased with Subtraction Difficulty. The interaction between Subtraction Difficulty and Text-Entry Difficulty also reached significance, which is due to the increased response times in the hard-hard condition, as was predicted. All other effects were not significant.

The lower panels of Figure 4.11 show the response times on the subtraction task. This is the time between clicking a button in the text-entry task and entering a digit in the subtraction task. Again, first responses of a trial were removed. The main effects of Subtraction Difficulty and Text-Entry Difficulty reached significance (see Table 4.10), showing an increase in response times for both effects. Response times increased even more when both tasks were hard, as shown by the significant interaction effect of Subtraction Difficulty and Text-Entry Difficulty. All other effects did not reach significance.

The two top panels of Figure 4.12 show the accuracy on the text-entry task. The main effects of Subtraction Difficulty and Text-Entry Difficulty reached significance, as did the interaction effect between Subtraction Difficulty and Text-Entry Difficulty: Accuracy decreased with the two main effects, and even more when both tasks were hard. The other effects did not reach significance.

The lower panels of Figure 4.12 show the accuracy on the subtraction task. The main effect of Subtraction Difficulty was significant; the interaction between Subtraction Difficulty and Text-Entry Difficulty showed a trend towards significance. The other tests did not reach significance.

Discussion

This experiment was performed to test why two results of the fMRI experiment were slightly different from previous experiments: First, the interaction effect in the response times of the subtraction task was absent in the fMRI experiment, and second, the pattern of response times of the text-entry task was different. The current experiment, with the same interface as the fMRI experiment, did find a significant interaction effect in the response times of the subtraction task, and shows a ‘normal’ pattern in the response times of the text-entry task. This suggests that the small differences in the behavioral data of the fMRI experiment are due to the changed environment and the low number of participants, not to the new interface.

The Neural Correlates of Problem States: Model-Based fMRI Analysis

In which we apply a novel model-based fMRI analysis method to locate the neural correlates of the problem state resource, declarative memory, and other resources.

This chapter was previously published as:

Borst, J.P., Taatgen, N.A., & Van Rijn, H. (2011). Using a Symbolic Process Model as input for Model-Based fMRI Analysis: Locating the Neural Correlates of Problem State Replacements. *NeuroImage, 58*(1), 137–147.

5

Chapter

Chapter 5

The Neural Correlates of Problem States: Model-Based fMRI Analysis

Abstract

In this paper, a model-based analysis method for fMRI is used with a high-level symbolic process model. Participants performed a triple-task in which intermediate task information needs to be updated frequently. Previous work has shown that the associated resource – the problem state resource – acts as a bottleneck in multitasking. The model-based method was used to locate the neural correlates of "problem state replacements". To analyze the fMRI data, we fit the computational process model to the behavioral data and regressed the model's activity against the fMRI data. The brain region responsible for the temporary representation of problem states, the inferior parietal lobule, and the brain region responsible for long-term storage of problem states, the inferior frontal gyrus were thus identified. These results show that model-based fMRI analyses can be performed using high-level symbolic cognitive models, enabling fine-grained exploratory fMRI research.

Introduction

If one wants to find the neural correlates of a theorized cognitive process using the classical fMRI analysis method of cognitive subtraction (e.g., Cabeza & Nyberg, 2000; Logothetis, 2008), the first step is to translate the theory into suitable experimental conditions. Then, an experimental condition placing demands on the process of interest is compared to a control condition. The control condition is the same as the experimental condition except for the absence of the process under investigation. Brain areas that are more active in the experimental condition than in the control condition are assumed to be involved in the cognitive process of interest (e.g., Friston, Ashburner, Kiebel, Nichols, & Penny, 2007). However, it would be better to localize cognitive functions in a more direct way. Especially for more complex tasks, the translation of theory into experimental conditions is non-trivial. In complex tasks, central cognitive processes are often used in all experimental conditions (although with a different frequency or temporal pattern), which makes it difficult to find a good control condition that does not include the process of interest. A way to address this problem and to localize brain functions in a more direct way is to use model-based fMRI analysis (e.g., Gläscher & O'Doherty, 2010; O'Doherty et al., 2007).

In model-based fMRI analysis, information coming from a computational model that simulates the process of interest is correlated against fMRI data, showing which brain areas show activation patterns that are consistent with the process of interest. This method has proven to be very successful in locating brain areas involved in reinforcement learning (e.g., Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Hampton, Bossaerts, & O'Doherty, 2006; Haruno & Kawato, 2006; Kim, Shimojo, & O'Doherty, 2006; Wunderlich, Rangel, & O'Doherty, 2009). Parameters of mathematical reinforcement models were correlated against brain data, showing which brain areas are involved in the reinforcement learning process. In this article we will use the model-based method with a higher-level symbolic cognitive model. Such a higher-level model not only simulates a particular process, but the whole task including, for example, visual and motor processes. Instead of correlating model parameters, we will correlate the presence and absence of activity of cognitive resources against brain data, showing where the cognitive resources are best represented in the brain. This way, we will investigate whether predictions derived from a high-level process model can be used for model-based fMRI, and whether this combination allows for more direct exploratory fMRI analyses.

The Problem State Resource

We will use model-based fMRI to analyze data of a relatively complex experimental paradigm, which was developed to investigate the neural correlates of the "problem state resource" (Borst, Taatgen, Stocco, et al., 2010). The problem state resource is defined as the part of working memory that is available at no time cost (Anderson, 2005), as opposed to other elements in working memory that take time to retrieve and use (e.g., McElree, 2001). It is normally used to represent intermediate information in a task, and can at most contain one chunk of information (Borst, Taatgen, & Van Rijn, 2010). Thus, the concept of a problem state resource is comparable to the focus of attention in working memory theories that pose an extremely limited focus of attention (e.g., Garavan, 1998; McElree, 2001). The concept of a central problem state resource originates from a series of neuroimaging experiments by Anderson and colleagues, who found that activity in the posterior parietal cortex correlates with the number of transformations of mental representations (Anderson, 2005; Anderson, Albert, et al., 2005; Anderson et al., 2003; Sohn et al., 2005).

Although Anderson and colleagues assumed on functional grounds that the problem state resource contains at most one chunk of information, we have recently provided empirical evidence for this assumption. In a series of experiments, we showed that the problem state resource is a source of interference when required by multiple tasks at the same time (Borst & Taatgen, 2007; Borst, Taatgen, & Van Rijn, 2010). A computational cognitive model was developed to account for the observed multitasking interference. The basic assumption of the model is that when multiple tasks require a problem state, the contents of the problem state resource have to be replaced on each switch between tasks. That is, on every alternation the problem state of the previous task is stored in declarative memory, while the problem state of the current task is recalled from declarative memory and restored to the problem state resource. The model incorporating these time-consuming and error-prone problem state replacements provided a good match with the interference effects in the data. The current experiment was performed to find the neural correlates of the resources that are used by the model, which are, apart from the problem state resource, associated with vision, manual action, and declarative memory.

Materials and Methods

Behavioral Experiment

To locate the neural correlates of the model's resources, we used a triple-task design in which participants alternated between solving subtraction problems and entering text, while performing a listening comprehension task simultaneously (Figure 5.1 shows a screenshot of the experiment). Both the subtraction task and the text-entry task had two versions: an easy version that did not require maintenance of a problem state and a hard version that did.

In the subtraction task participants had to solve 10-column subtraction problems. Although participants were shown only one column at a time to minimize eye and head movements, participants were trained to perceive these columns as part of a full 10-column subtraction problem. In the easy version the upper term was always larger or equal to the lower term: no carrying was required. However, in the hard version participants had to carry in 6 out of the 10 columns; thus, participants had to remember whether a carry was in progress while performing the text-entry task.

In the text-entry task participants had to enter 10-letter strings. In the easy version a single letter was shown, which the participants had to enter. After solving one column of

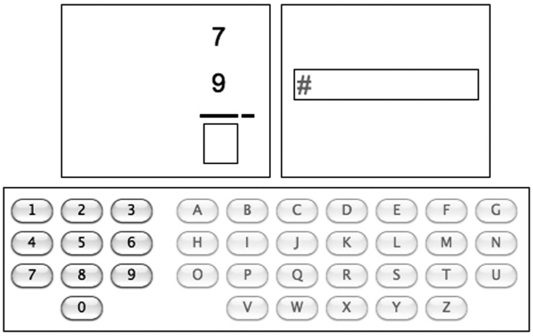


Figure 5.1 Screenshot of the experiment.

the subtraction problem, a new letter was shown, etc. In the hard version, a complete 10-letter word was shown once at the start of a trial, but as soon as the participant entered the first letter, the word disappeared and had to be entered letter by letter without feedback. Thus, in the hard version of the text-entry task participants had to remember what word they were entering.

Because participants had to alternate between the tasks after every number and letter, they had to keep track of whether a carry was in progress and what word they were entering (and the position within the word) in the hard versions of the tasks while giving a response on the other task. Supported by results from previous experiments (Borst, Taatgen, & Van Rijn, 2010), we assumed that participants used their problem state resource to keep track of the absence or presence of a carry and the words, but did not use this resource in the easy conditions. In half of the trials participants also had to perform a listening comprehension task. As the listening task is not the focus of the current article, we collapsed over this task if not mentioned otherwise. A detailed description of the Methods and discussion of the setup can be found in Borst, Taatgen, Stocco, et al. (2010).

fMRI Procedures and Analysis

The fMRI data were collected with a Siemens 3T Allegra Scanner using a standard radio frequency head coil. Each functional volume existed of 34 axial slices (3.2 mm thickness, 64 × 64 matrix, 3.125 × 3.125 mm per voxel), acquired using echo-planar imaging (2000 ms TR, 30 ms TE, 79° flip angle, 200 mm field of view, 0 slice gap, with AC-PC on the 11^{th} slice from the bottom). Functional acquisition was event-related; scanning onset was synchronized with stimulus onset. Anatomical images were acquired using a T1-weighted spin-echo pulse sequence at the same location as the functional images but with a finer resolution (3.2 mm thickness, 200 mm field of view, 256 × 256 matrix, 0.78125 × 0.78125 mm in-plane resolution).

The data were analyzed using SPM5 (Wellcome Trust Centre for Neuroimaging, London). This included realigning the functional images, coregistering them with the structural images, normalizing the images to the MNI (Montreal Neurological

Institute) ICBM 152 template, and smoothing them with an 8 mm FWHM Gaussian kernel.

Participants

Thirteen students of Carnegie Mellon University participated in the experiment. The data of three participants were excluded (one participant fell asleep in the MRI scanner, one ignored the listening task, and with one fMRI recording problems were encountered) leaving 10 complete data sets (3 women, average age 21.9, range 19–28, right-handed). All participants had normal or corrected-to-normal vision and normal hearing. Informed consent as approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh was obtained before the experiment. Participants received US$ 100 compensation.

Results

Behavioral Results

All reported *F*- and *p*-values are from repeated measure analyses of variance (ANOVAs), effects were judged significant when a .05 significance level was reached, and accuracy data were transformed using an arcsine transformation before performing ANOVAs. Outliers in response times were eliminated by means of a two-step procedure. First, response times faster than 250 ms and slower than 10,000 ms were removed. Then, data exceeding 3 standard deviations from the mean per condition per participant were excluded. Overall, 2.4% of the data was discarded. First responses on both tasks were removed per trial.

As the listening task is not the focus of the current paper, we collapsed over levels of difficulty for this task if not mentioned otherwise.[1] Response time on the text-entry task was defined as the time between entering a digit in the subtraction task and entering a letter in the text-entry task, while a response time on the subtraction task was defined as the time between entering a letter in the text-entry task and entering a digit in the subtraction task. Figure 5.2(a) shows the response times on the text-entry task (left) and the subtraction task (right); Table 5.1 contains the results of the ANOVAs. The interaction between Subtraction Difficulty and Text-Entry Difficulty was significant in the text-entry task, as were both main effects. Interestingly, response times decreased when the text-entry task was hard but increased when the subtraction task was hard. We have come across this effect before (e.g., Experiment 2 in Borst, Taatgen, & Van Rijn, 2010). It can be explained by assuming that in the hard condition participants know what word they are entering and thus do not need any additional visual input, but in the easy condition participants first have to look at the screen to see which letter they have to enter next. Our computational model (Borst, Taatgen, & Van Rijn, 2010) fitted these results. In the subtraction task the interaction effect failed to

[1] *Please note that the listening task has hardly any influence on the subtraction and text-entry results (Borst, Taatgen, Stocco, et al., 2010; Borst, Taatgen, & Van Rijn, 2010).*

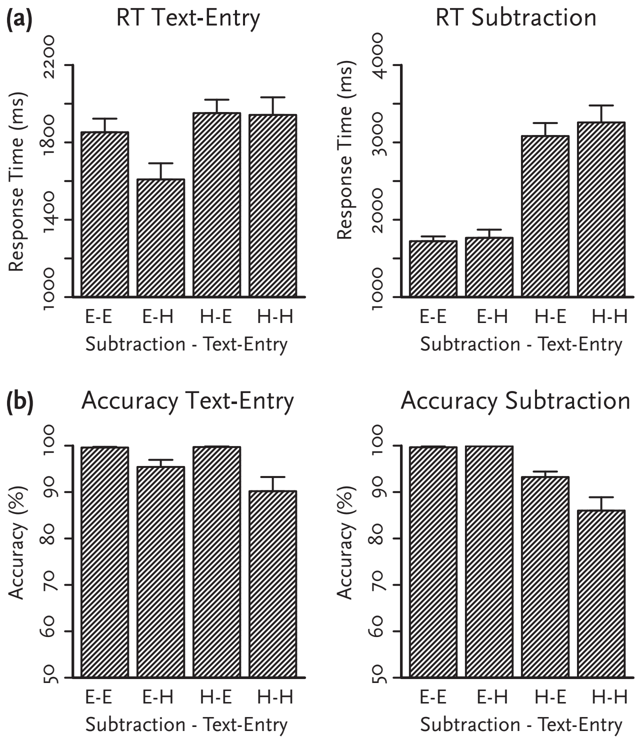


Figure 5.2 Behavioral results. RT = response time, E/E = easy–easy, E/H = easy–hard, etc., error bars indicate standard error.

reach significance, but both main effects did: a small increase of response times when text entry was hard, and a large increase when the subtraction task was hard.

Figure 5.2(b) shows the accuracies, in which the interaction effects between Subtraction Difficulty and Text-Entry Difficulty reached significance for both tasks. The main effects of the tasks were also significant: Subtraction Difficulty for the subtraction task and Text-Entry Difficulty for the text entry task.

Based on similar effects on response times and accuracy we previously argued that the results of this type of experiment support the idea of a problem state bottleneck (Borst, Taatgen, & Van Rijn, 2010). In the easy tasks, no intermediate results need to be stored in the problem state resource. If a task is hard, accurate performance on a task requires storing intermediate results. Although participants had to alternate between tasks, the combination of one hard and one easy task is not problematic, since the problem state resource is not overwritten during the easy task. If, however, both tasks are hard, both tasks require the use of the problem state resource. Therefore, on each step in the task in the *hard–hard* condition the problem state resource has to be swapped out: on each step an old problem state is retrieved from declarative memory and restored to the problem state resource, overwriting the problem state of the other

Table 5.1 ANOVA results of the behavioral data. Interaction is the interaction between Subtraction Difficulty and Text-Entry Difficulty.

	Response Times			Accuracy		
Source	$F(1,9)$	p	η_p^2	$F(1,9)$	p	η_p^2
			Text-Entry Task			
Subtraction	32.13	<.001	.78	5.20	.049	.37
Text-Entry	5.62	.042	.38	44.27	<.001	.83
Interaction	12.10	.007	.57	6.38	.03	.41
			Subtraction Task			
Subtraction	83.94	<.001	.90	96.33	<.001	.91
Text-Entry	5.36	.046	.37	4.04	.075	.31
Interaction	2.66	.137	.23	21.16	.001	.70

task. This results in the typically observed over-additive interaction effects.[2] To test whether this effect can indeed explain the observed data, we developed a computational cognitive model, which we will discuss in the next section.

Cognitive Model

To account for the data, we used a high-level symbolic cognitive model (Borst, Taatgen, & Van Rijn, 2010), developed in the cognitive architecture ACT-R (e.g., Anderson, 2007). Most important for the task at hand are the problem state resource and resources associated with vision, manual actions, and declarative memory. The model uses the visual resource to perceive the stimuli; this resource is assumed to do focused processing of attended stimuli. The manual resource was used to make responses; it operates the "hands" of the model. The declarative memory resource was used to retrieve facts (e.g., "5 - 2 = 3"). Facts in ACT-R have a certain activation level, which represents frequency and recency of use (e.g., Anderson & Schooler, 1991). This activation level determines the probability of retrieving a fact, and the speed with which a fact is retrieved. For example, a simple subtraction fact such as "5 - 2" is probably used very often, and has therefore a high activation level. In contrast, "17 - 8" will have been used less often in the past, and will therefore have a lower activation level and take a little more time to retrieve.

The problem state resource is used to maintain intermediate information and is therefore of particular interest for the current article. Information in the problem state resource can be accessed at no time cost, but it takes 200 ms to replace it (e.g., Anderson, 2005). Problem states that are discarded from the problem state resource are still available in declarative memory, and can be retrieved and restored later.

[2] *That the interaction for the response times of the subtraction task in the current experiment failed to reach significance is probably a power issue, as previously reported experiments showed significant effects (Borst, Taatgen, & Van Rijn, 2010). Moreover, an extensive behavioral pilot experiment with the exact same experimental setup as the current experiment also showed both interactions. We report the results of this pilot experiment in the Appendix.*

Using a cognitive architecture makes it meaningful to model all components relevant for a task such as visual, manual, and declarative memory processes (e.g., Cooper, 2007; Newell, 1990): the architecture provides the time it takes to move the mouse or retrieve a fact from memory, and as such the time courses of when the different resources are used. These elements of the architecture have received extensive experimental support (e.g., Anderson, 2007 and see http://act-r.psy.cmu.edu/).

To account for multitasking aspects the model uses threaded cognition theory (Salvucci & Taatgen, 2008, 2011; Salvucci, Taatgen, et al., 2009). According to threaded cognition theory, multiple tasks can be active concurrently, but a cognitive resource can only be used by one task at a time. Thus, the problem state resource can only maintain information for a single task. However, in the *hard–hard* condition of the present experiment, a problem state is required for both tasks. The problem state then has to be replaced on each step of a trial (participants had to alternate between the subtraction and text-entry tasks), which takes time (a declarative retrieval and 200 ms problem state restoration, see Borst, Taatgen, & Van Rijn, 2010 for details). In contrast, in the *easy–easy* condition the problem state resource is not used at all, while in the *easy–hard* and *hard–easy* conditions it is only used for one task. Because problem states have to be restored and replaced at each step in the *hard–hard* condition, this leads to an over-additive interaction effect on response times. Furthermore, because the model sometimes retrieves an incorrect problem state from declarative memory, it also gives an explanation for the interaction effect on accuracy.

Previously, the model was fitted to data of Experiment 1 in Borst, Taatgen, and Van Rijn (2010, p. 370), and shown to give a good account of the data (R^2 for response times and accuracy approached 1). Subsequently, the model was used to predict the data of Experiment 2 and 3 in the same paper, showing that the model was capable of generalizing to different data sets. Experiment 3 in that paper also included the listening task; the model fit well to the data of all three tasks in that data set. We use the exact same model in the current paper to analyze the fMRI data. Thus, the model takes the small influence of the listening task on the timing of the other tasks into account. The listening task itself, at least for the purposes of the current model, only requires use of the declarative memory resource, and does not use any of the other resources. However, we analyzed declarative memory only in the non-listening condition. An extensive description of the (lack of) influence of listening can be found in Borst, Taatgen, and Van Rijn, 2010.

Model-Based fMRI Analysis – Method

We will now turn to the model-based fMRI analysis data to locate the model's resources. For the classical fMRI analysis method of cognitive subtraction, one typically defines stimulus functions that correspond to the experimental conditions (e.g., Friston et al., 2007). These stimulus functions are entered into a general linear model (GLM), which shows brain areas in which activity correlates with the conditions of the experiment. Stimulus functions used for classical fMRI analyses are coarse in the sense that they assume stable activation during a complete trial – an assumption that often does

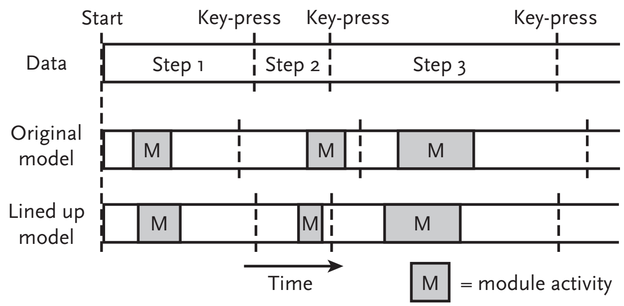


Figure 5.3 Demonstration of the linear transformation that was used to line up the model data with the participants' data.

not hold. During a trial in the current experiment, for example, participants solve a 10-column subtraction problem and enter a 10-letter word. These processes involve multiple fixations, manual actions, memory retrievals, and continuous maintenance and updating of problem states, and can thus not be characterized by constant activation throughout a trial. To construct more detailed stimulus functions, we fitted the cognitive model to the data, and entered the activity of the model's resources as stimulus functions into the general linear model. This method takes into account when and how often a resource is used during a trial, instead of assuming constant activation.

To approximate the cognitive processes at trial level, we ran the model for each participant on the same stimuli as the participant received, in the same order, including all non-experimental components, such as fixation and feedback screens. To further improve the timing of the model, we lined up the model's responses with the participant's responses. Figure 5.3 gives a schematic overview of the procedure. The first line represents data, with key-presses as dashed lines. The second line shows a model simulation of these steps. As the model is regressed directly against brain data it is important to have a correct time mapping between model and data (Gläscher & O'Doherty, 2010): it does not make sense to compare a fixation in the model to a key-press in the data. Therefore, we used a linear transformation to line up the key-presses of the model to the key-presses of the participants. Line 3 in Figure 5.3 shows the result: the transformation causes Step 1 of the model to increase a little in length and Step 2 to decrease in length. Not only the key-presses of the model are shifted, but also cognitive resource activity within a step is shifted and in- or decreased in length (represented by the grey boxes in Figure 5.3). The resulting activity for four cognitive resources during four different trials in the experiment is shown as grey lines in Figure 5.4.

As the next step, the stimulus functions were convolved with a hemodynamic response function. The convolved stimulus functions are displayed in black in Figure 5.4. For model-based fMRI analysis it is crucial that the different resources of the model make different predictions, because otherwise different resources cannot be distinguished. Figure 5.4 shows clearly different patterns between the problem state resource and

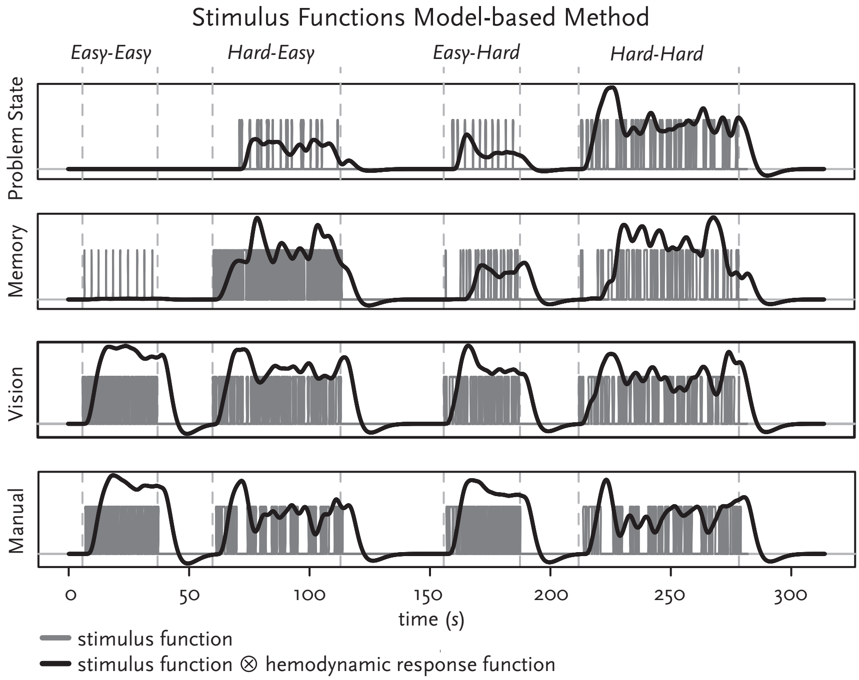


Figure 5.4 Stimulus functions and convolved stimulus functions for the model-based analysis method, for four cognitive resources: the problem state resource, declarative memory, vision, and the manual motor resource. Easy–Hard etc. = Easy Subtraction – Hard Text-Entry.

declarative memory[3] on the one hand, and the visual and manual resources on the other hand. It is important to note that problem state activity of the model relates to changing the contents of the problem state resource, not to passive maintenance of information. The problem state resource and declarative memory are used most in the *hard–hard* condition because of repeated problem states replacements, while they are used less in the easier conditions. The visual and manual resources show a different pattern: they are used for roughly the same amount of time in all conditions, but their activity is spread over a larger period of time in the harder conditions, resulting in a lower BOLD prediction (the same amount of visual information has to be processed and the same amount of key-presses have to be made, but as response times are higher more time is available in the more difficult conditions). The problem state resource and declarative memory can also be distinguished from each other: declarative memory is used often in the *hard subtraction – easy text-entry* condition (many subtraction facts have to be retrieved from memory to process the carries), resulting in high BOLD predictions. The problem state resource only shows intermediate BOLD levels in this condition (see Figure 5.4). The visual and manual stimulus functions, on the other

[3]*The fact that there is hardly any BOLD response predicted in the easy–easy condition of the declarative memory resource is caused by the very short (~ 10 ms) retrievals in that condition (only subtraction facts under 10 have to be retrieved, e.g., 5 - 2 = 3).*

Table 5.2 Correlations between the different regions and predictions. Please note that correlations with declarative memory were calculated on the non-listening data, because that was the data that was used for the model-based analysis.

Correlation between	Prediction	Data
Problem State – Declarative Memory	.69	.65
Problem State – Vision	.51	.23
Problem State – Manual	.43	.44
Declarative Memory – Vision	.22	.35
Declarative Memory – Manual	.21	.62
Vision – Manual	.93	.38

hand, are very similar. Table 5.2 shows the inter-correlations between the stimulus functions (the prediction column shows the correlations between the predictions). The correlation between the visual and the manual resource is .93, which makes it difficult to identify separate areas for these resources.

These stimulus functions were then entered one by one into a GLM to see which brain areas correlated with the predicted activity of the resources. Each stimulus function was accompanied by its 'opposite': a function showing when the resource was not active.[4] Feedback screens and the screens indicating conditions were also entered into the GLM; fixation screens formed the baseline. Contrast images were made for the individual participants, and entered into second level random-effect group analyses.

Model-Based fMRI Analysis – Results

Figure 5.5 shows the results for (a) the problem state resource, (b) declarative memory, (c) vision, and (d) the manual resource. The column on the left shows the regions that were identified by the model-based fMRI analysis. A threshold of $p < .01$ (FWE-corrected) and 100 contiguous voxels was applied to the results, with the exception of declarative memory, for which a threshold of .05 was used (FWE-corrected; see below for an explanation). Crosshairs in Figure 5.5 are located at the most significant voxel, except for the manual resource and the visual resource (see below). The *xyz*-coordinates indicate the most significant voxel in MNI-coordinates in the located region. The white squares show the existing mapping between ACT-R's resources and brain regions (e.g., Anderson, 2007), which can be used for confirmatory analyses (e.g., Anderson et al., 2008; Borst, Taatgen, Stocco, et al., 2010).

The area that corresponded best to problem state activity was located in the inferior parietal lobule, around the intraparietal sulcus. Declarative memory also showed activation in that area, but the best fitting area was located around the inferior

[4] *While the 'non-activity' of a resource is implicitly modeled by its stimulus function (1 = resource active, 0 = not active), we added a non-activity regressor to distinguish it from the fixation screens. Because non-activity often takes place in between activity of a resource, due to the convolution with the HRF there is usually some activity present on these scans, unlike on the fixation scans. To account for this, we added 'non-activity' as a separate regressor.*

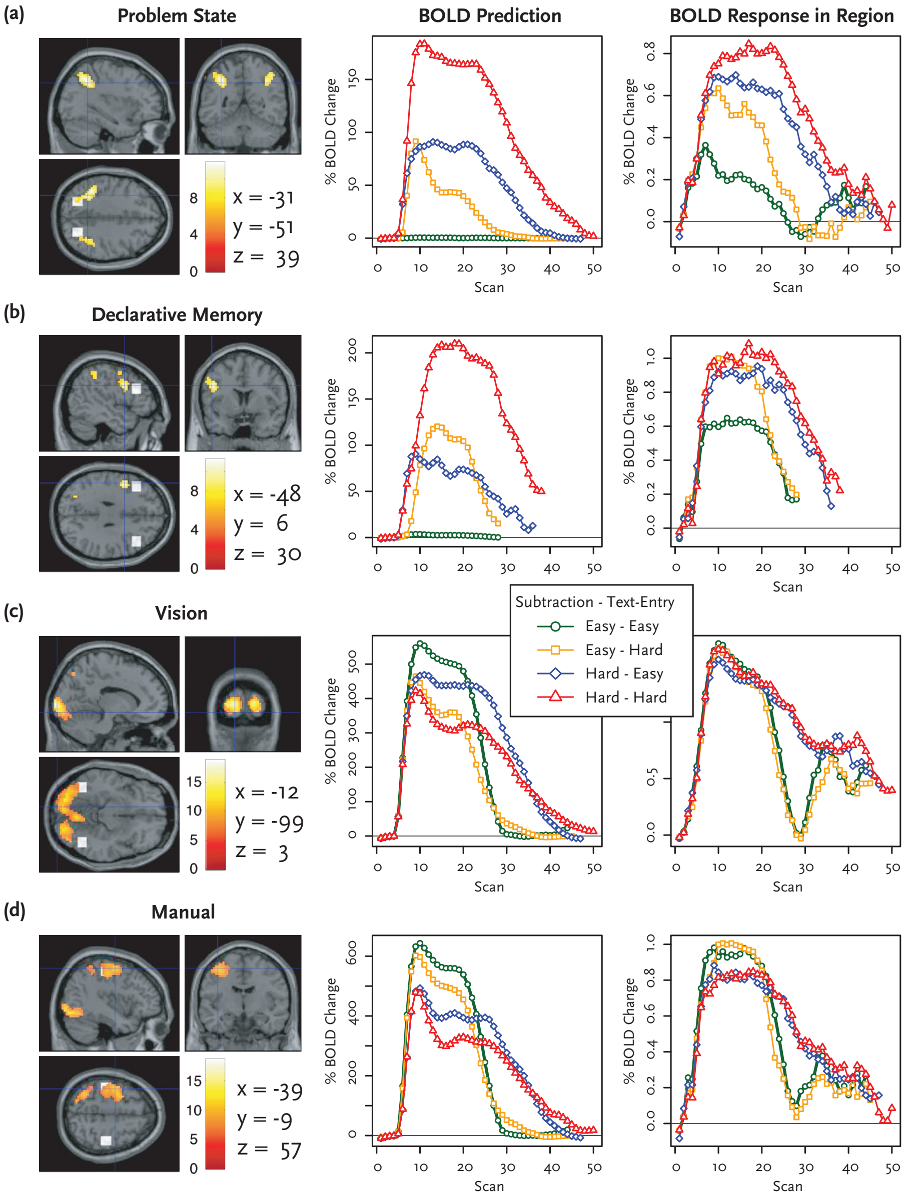


Figure 5.5 Results of the model-based analysis method, for (a) the problem state resource, (b) declarative memory, (c) vision, and (d) the manual motor resource. (a), (c), and (d) with a Family Wise Error threshold of $p < .01$ and 100 contiguous voxels, (b) with an FWE threshold of $p < .05$ and 100 contiguous voxels. White squares represent predefined mappings between ACT-R's resources and the brain. Crosshairs are centered at the most significant voxel, except for the manual, which is centered on the most significant voxel in the cluster in the motor cortex and for the visual, which was moved down 13 mm to enable comparison to the predefined region of ACT-R. The middle column shows the stimulus function that was entered into the GLM, averaged per condition and trial, the right column the measured BOLD response in the located area.

frontal gyrus. The threshold for declarative memory was increased to $p < .05$.[5] This was necessary due to a more limited data set, as we only used the trials in which the listening task was not present. If we included the trials with the listening task, the best fitting area for declarative memory coincided with the aural regions, because the model is not able to separate auditory processing of the incoming speech and the subsequent updating of declarative memory.

Figure 5.5(c) and (d) show the results of the visual and manual resources. As discussed above, the stimulus functions were very similar as moving the mouse in the model is almost always accompanied by moving the eyes, and this was reinforced by the convolution with the HRF. It was therefore not surprising that the analyses yielded very similar results. The most significant area for both was the occipital visual area, but both also showed a fitting area in the motor cortex. This area was a little larger for the manual resource than for the visual resource, and as we know that manual actions are represented in the motor cortex, we centered the results of the manual resource on that area. The crosshairs of the visual area were moved down 13 mm, to enable comparison with ACT-R's predefined region (however, the coordinates indicate the most significant voxel).

Table 5.3 ANOVA results of the area under the curve in the located regions. "Interaction" is the interaction between Subtraction Difficulty and Text-Entry Difficulty.

Source	*F*(1,9)	*p*	η_p^2
	Problem State		
Subtraction	121.24	< .001	.93
Text-Entry	23.07	< .001	.72
Interaction	< 1	-	-
	Declarative Memory		
Subtraction	52.25	< .001	.85
Text-Entry	15.76	.003	.64
Interaction	< 1	-	-
	Vision		
Subtraction	43.51	< .001	.83
Text-Entry	< 1	-	-
Interaction	4.66	.059	.34
	Manual		
Subtraction	6.81	.028	.43
Text-Entry	< 1	-	-
Interaction	2.06	.185	.19

The middle column of Figure 5.5 shows the stimulus functions that were entered into the GLM, averaged per condition per trial. Thus, what is shown here is the activity of the model convolved with the BOLD response over the course of entering 10 numbers in the subtraction task and entering 10 letters in the text-entry task. The x-axis represents time in scans (1 scan = 2 seconds). The y-axis shows % BOLD change (the height of the curves is not important for the GLM, only the relative magnitude of the curves). The right column of Figure 5.5 shows the measured BOLD response in the 100 most significant voxels for the located regions. ANOVA results of the area under the curve (reflecting total activation in a trial, see e.g., Anderson, 2005; Stocco & Anderson, 2008) are reported in Table 5.3. In general, the graphs show that the model-based fMRI method is able to identify patterns of activation in the brain that are very similar to the predictions of the model. Where for the problem state and declarative memory resources the *hard–hard* condition shows most activation, this is

[5] *If we decrease the p-values to .001 the same areas are found for all cognitive resources (except that the number of consecutive voxels had to be lowered to 25 for declarative memory).*

not the case for the manual and visual resources[6], as predicted by the model. On the other hand, the fit is not perfect. For instance, while the model predicted no activity at all in the *easy–easy* condition for the problem state resource, this was not found in the located region. This indicates that the model-based analysis is not limited to identifying regions with a perfect fit, but locates regions that correlate significantly with the model predictions.

Discussion

The model-based method was able to find neural correlates corresponding to the cognitive resources of the model. Instead of using the experimental conditions as a basis for the analysis, the model-based method allows for assessing directly which parts of the brain correlate significantly with model predictions. In this paper we have shown that this is possible with a detailed high-level symbolic cognitive model (as compared to the more low-level mathematical models that have been used previously; e.g., Daw et al., 2006; Gläscher & O'Doherty, 2010; Hampton et al., 2006; Haruno & Kawato, 2006; Kim et al., 2006; O'Doherty et al., 2007; Wunderlich et al., 2009). Instead of focusing on one process, using a high-level model integrated in a cognitive architecture allows for analyzing all cognitive processes that are involved in the task, and thus localizing multiple resources in one experiment. Furthermore, because the analysis is based on a cognitive model, the fMRI results are grounded in the theoretical framework the model was built on, providing a functional explanation of the results (Gläscher & O'Doherty, 2010; O'Doherty et al., 2007).

The experiment was set up to test a hypothesis related to the problem state resource. Because activity in the model is related to changes of the problem state resource, the located region represents these changes, and not storage of problem states per se. The analysis showed that the model's problem state activity corresponded best to a region focused in the inferior parietal lobule, but also included parts of the superior parietal lobule and the intraparietal sulcus. The intraparietal sulcus has been linked previously to ACT-R's problem state resource (e.g., Anderson, 2005; Anderson, Albert, et al., 2005; Anderson et al., 2003; Sohn et al., 2005). While it is most often referred to in connection with spatial working memory, it is also known to be involved in object and verbal working memory (e.g., LaBar et al., 1999; Smith et al., 1998; Wager & Smith, 2003), functions that are attributed to the problem state resource.

Use of declarative memory also correlated with activity in the identified problem state region, but the best fitting area was the inferior frontal gyrus, slightly anterior to the standard ACT-R region for declarative memory (e.g., Anderson, 2007; Anderson et al., 2008). This region is known to be involved in memory retrieval (e.g, Cabeza, Dolcos, Graham, & Nyberg, 2002; Fletcher & Henson, 2001; Wagner, Maril, Bjork, & Schacter, 2001). More interestingly, if we lower the significance threshold, it becomes clear that both areas are part of a larger fronto-parietal network, a network that is often

[6] *The graph for declarative memory ends earlier than the graphs of the other resources because only trials without the listening task are taken into account; these trials are shorter than the trials with the listening task.*

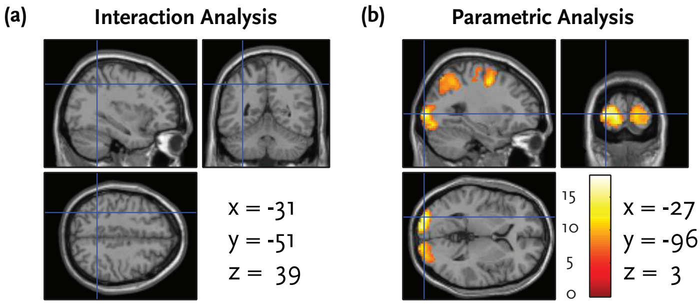


Figure 5.6 Results of traditional fMRI analysis methods, for (a) a classical interaction analysis, and (b) a parametric modulation analysis (see the main text for details). The results were thresholded with (a) FWE $p < .05$ and (b) FWE $p < .01$ and 100 contiguous voxels.

implicated in working memory research (e.g., Collette, Hogge, Salmon, & Van der Linden, 2006; Collette & Van der Linden, 2002). While both declarative memory and problem state activity are associated with functions of working memory, the current analysis implies that retrieving information is done via an area around the inferior frontal gyrus, while maintaining and updating working memory is performed in the parietal regions.

Model-Based fMRI versus Traditional Analysis Methods

We have shown that model-based fMRI makes it possible to directly locate the neural correlates of model resources, and therefore allows for fine-grained exploratory fMRI analyses. However, does it perform better than traditional methods? While there is no traditional method that allows for direct localization of model components, we used two existing methods to localize the over-additive interaction effect that was predicted by the model's problem state resource (see the BOLD prediction in Figure 5.5(a), middle column), and compared the results of these methods to the results of the model-based method. First, we used a classical cognitive subtraction approach to find an interaction effect; when that failed we tried a parametric method that is somewhat similar to the model-based method.

For the classical cognitive subtraction analysis we defined a stimulus function for each condition in the experiment. Subsequently, the four stimulus functions were convolved with a hemodynamic response function and entered into a general linear model. We then tested for an over-additive interaction effect of Subtraction and Text-Entry Difficulty by contrasting the difference between the *hard–hard* condition and the *hard–easy* condition against the difference between the *easy–hard* and the *easy–easy* condition (i.e. (*hard–hard* - *hard–easy*) - (*easy–hard* - *easy–easy*); e.g., Friston et al., 2007). The results are shown in Figure 5.6(a): no voxels crossed the FWE significance threshold of .05. Thus, the traditional cognitive subtraction method was unable to find a region that showed the predicted interaction effect, and could not be used to locate the neural correlates of the problem state resource.

We then used a method that is more similar to model-based fMRI: a parametric analysis (e.g., Büchel, Wise, Mummery, Poline, & Friston, 1996; M. S. Cohen, 1997). For this method we defined one stimulus function for all conditions in the experiment, and then specified a parametric model, with 0 for the *easy–easy* condition, 1 for the *hard–easy* and *easy–hard* conditions, and 3 for the *hard–hard* condition. This method is comparable to the model-based method, except that the amplitudes of the different conditions have to be specified by the researcher, instead of the model providing these estimates (note that the model also predicts a detailed pattern within each trial and differences between participants, see the next section). The results are shown in Figure 5.6(b): while the problem state region was found, visual and motor areas also showed significant activation, with the most significant voxel being located in the visual cortex. Thus, the parametric method yielded less specific results than the model-based method, and, moreover, indicated a seemingly incorrect region (compared to previous results, e.g., Anderson et al., 2008).

Model-based fMRI thus outperformed these two traditional methods. There are obviously other possibilities to analyze the data, but, to our knowledge, none of these options can directly show the neural correlates of resources in a model. Additionally, model-based fMRI allows for locating multiple resources in one experiment, without having to adapt the experimental design for each resource, which is the case for traditional methods. In the next section we performed a detailed analysis to investigate what enabled the model-based fMRI analysis to locate the model's resources.

What makes the Model-Based Method Powerful?

To arrive at the reported results, the model-based method uses a computational cognitive model to look in a more informed way at fMRI data. The model-based stimulus functions not only contain differences between conditions (as in classical fMRI analyses), but also differences per participant and trial, and even a detailed temporal pattern within each trial (Figure 5.4). To assess what drives the results, we compared a series of models that incorporate increasing levels of detail. First, we constructed four different stimulus functions for each cognitive resource (Figure 5.7). Please note that we used stimulus function 5.7(d) for the model-based analysis described above; (a)-(c) are only used for investigating which level of detail drives the results. The first stimulus function, 5.7(a), contains differences per condition, but was the same for all trials in a condition and all participants. The second stimulus function, 5.7(b), also contained differences per participant[7], while the third and the fourth stimulus functions in addition contained effects of trial[8]. The fourth stimulus function differed from the third with respect to

[7] *Effects of participant are caused by differences in speed between participants: some participants are slower in, for example, the easy–easy condition, resulting in a lower BOLD prediction (resource activity is spread more over the trial). However, all trials of the same condition of one participant have the same predicted BOLD amplitude.*

[8] *Effects of trial originate in the response times on a particular trial: for instance, for a quick trial with the same number of key-presses as a slow trial, a higher BOLD response is predicted for the manual motor resource.*

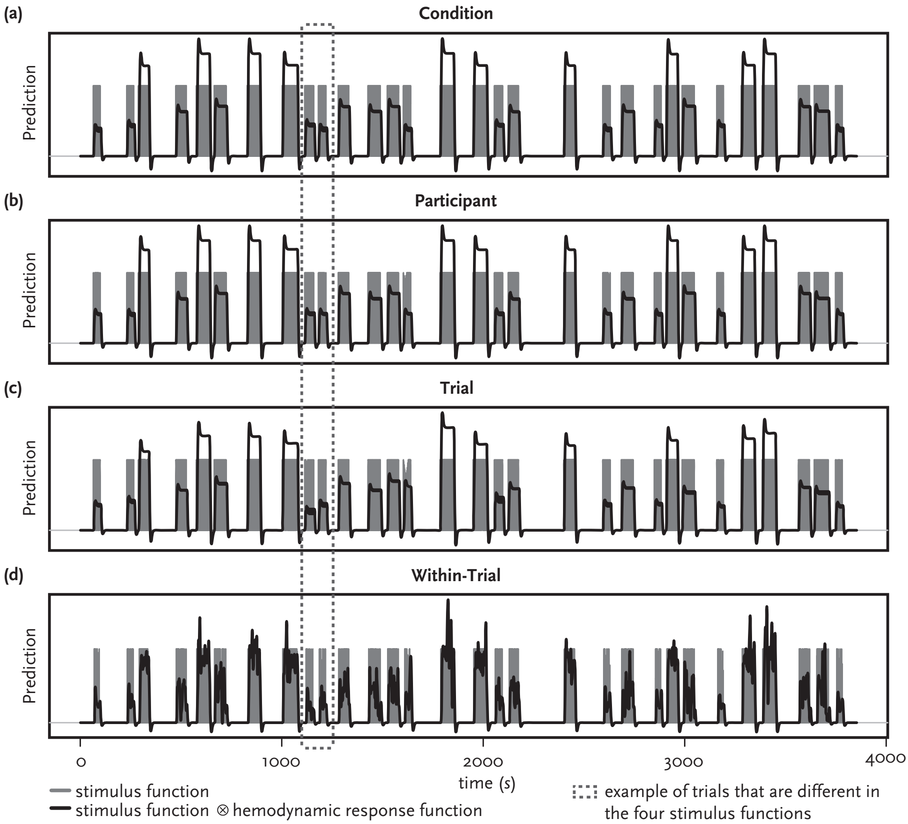


Figure 5.7 Example of four different stimulus functions, shown for one session of one participant for the problem state resource. These stimulus functions were used to test which properties of the stimulus functions were important for the model-based analysis. Stimulus function (a) only contains effects of condition, (b) of condition and participant, (c) of condition, participant, and trial, and (d) of condition, participant, trial, and of cognitive resource usage within a trial. While it is clear that (d) is different from (a)-(c), the differences between (a)-(c) are minor.

the temporal detail within a trial: the first three stimulus functions are smooth, while the fourth has a very detailed temporal structure[9] (i.e., this is the stimulus function that was used for the model-based analysis reported earlier in this paper). Figure 5.7 illustrates the four different stimulus functions (note that here a stimulus function is shown for one resource, with different levels of detail; in contrast, Figure 5.4 shows four stimulus functions for four different resources).

These stimulus functions were then entered one by one into linear-mixed-effects models (e.g., Baayen et al., 2008) with the stimulus function as a fixed effect and participant as a random effect. For each cognitive resource we constructed two LMEs,

[9] *This temporal pattern stems from the time course of cognitive resource usage within a trial, as explained in the 'fMRI – Model-Based Method' section.*

Table 5.4 Log-likelihood of linear-mixed-effects models indicating which properties of the stimulus functions are important in the model-based analysis, for the best fitting voxel per cognitive resource. Each stimulus function is more detailed than the previous one, e.g., the stimulus function 'Participant' also includes effects of condition. See the main text for details.

Stimulus Function	Problem State	Declarative Memory	Vision	Manual
Condition	-36,743	-14,463	**-44,476**	-33,163
Participant	**-36,733**	-14,459	-44,480	-33,139
Trial	-36,771	**-14,444**	-44,492	**-33,138**
Within-Trial	-37,029	-14,701	-44,848	-33,676

Table 5.5 Log-likelihood of linear-mixed-effects models indicating which properties of the stimulus functions are important in the model-based analysis, for the average of the 100 best fitting voxels per cognitive resource. Each stimulus function is more detailed than the previous one, e.g., the stimulus function 'Participant' also includes effects of condition. See the main text for details.

Stimulus Function	Problem State	Declarative Memory	Vision	Manual
Condition	**-28,398**	-11,537	**-39,221**	-29,227
Participant	-28,402	-11,534	-39,231	-29,214
Trial	-28,426	**-11,516**	-39,257	**-29,209**
Within-Trial	-28,823	-11,849	-39,736	-29,650

one to fit the stimulus function to the BOLD response in the best fitting voxel, and one to fit the average BOLD response in the 100 best fitting voxels. All stimulus functions correlated significantly with the data. The results, in the form of the log-likelihood of the fit, are listed in Table 5.4 (best fitting voxel) and Table 5.5 (average of the 100 best fitting voxels). As can be seen in these tables, for none of the resources was it helpful to include the temporal pattern within a trial. For declarative memory and the manual resource the model made correct predictions on a trial-by-trial level, for the problem state resource and vision on a condition level (for the 100 best fitting voxels; for the problem state resource on a participant level for the best fitting voxel). Based on this we can conclude that it is useful to include trial-by-trial differences in the model-based stimulus functions, but not the temporal pattern within a trial (either because the model predictions are not precise enough within a trial, or because the data is too noisy). Thus, the strength of the model-based analysis lies in the predicted amplitude levels per trial. By providing the analysis with these precise a priori estimates of the amplitudes of the BOLD response, the model allows for a better identification of the regions involved in the task than is possible with classical fMRI analysis methods.

ACT-R

While the model-based analysis can be used with different kinds of models, the current article focuses on a model implemented in the ACT-R architecture. ACT-R only predicts when and for how long a cognitive resource is used, and not the intensity with which a resource is used. For instance, it is conceivable that an automatized movement takes as much time as a novel movement, but less effort. While this can be seen as a shortcoming, accounting for intensity would introduce extra free parameters, weakening the current predictive power of the model, as we would have to fit them post-hoc. On the other hand, if those new parameters were to explain a significant portion of the variance in the experimental data, they would increase the generalizability of the model (see e.g., Pitt & Myung, 2002).

If we compare the model-based fMRI method to the standard method of relating ACT-R models to neuroimaging data using predefined regions (Anderson, 2007; Anderson et al., 2008; see Borst, Taatgen, Stocco, et al., 2010 for a region-of-interest analysis of the current task), two things become clear: First, the areas that were located with the model-based analysis are very close to the predefined regions normally associated with ACT-R, which are shown as white squares in Figure 5.5. Only the visual area is different: the located region overlaps with V1, while the predefined ACT-R region is located in the fusiform gyrus. Thus, it seems that lower level vision actually fits better with the ACT-R predictions than the slightly higher-level visual processing of the fusiform gyrus. Second, the strength of the model-based method lies in its exploratory nature. Using this method, we can not only validate cognitive models, but also determine which brain regions are involved in complex tasks.

Conclusion

In this paper we have shown that the technique of model-based fMRI can be used in combination with a high-level symbolic process model. The model-based analysis method uses the results of a computational cognitive model to look in a more informed way at fMRI data: it shows areas in the brain that correlate with activity of the model. This method is especially useful for cognitive functions that are hard to discriminate in a pure subtraction-based design, for instance working memory storage and updating: These processes go hand-in-hand, which makes it difficult to find experimental conditions with one process but without the other. However, when a good model is available, these processes would yield different stimulus functions, which could in turn lead to different regions in the fMRI analysis, or at least different focal points in networks of activity. Because the model-based analysis works by refining the stimulus function, the method can be used with all modeling techniques that yield information that is more detailed than the condition structure of an experiment, which is used as the stimulus function in classical fMRI analyses.

Appendix: Behavioral Results outside the Scanner

Here we report the results of the pilot experiment that we ran outside the fMRI scanner. This experiment has exactly the same setup as the experiment that is reported in the main text.

Participants

Twenty students of Carnegie Mellon University participated in the experiment (11 women, average age 20.6, range 18–23). All participants had normal or corrected-to-normal vision and normal hearing. Informed consent as approved by the Institutional Review Boards at Carnegie Mellon University and the University of Pittsburgh was obtained before the experiment. Participants received US$ 10 for performing the experiment.

Results

Outliers in reaction times were eliminated by means of a two-step procedure. First, response times faster than 250 ms and slower than 10,000 ms were removed. Then, data exceeding 3 standard deviations from the mean per condition per participant were excluded. Overall, 2.0% of the data was discarded. Accuracy data were transformed using an arcsine transformation before performing ANOVAs.

Figure 5.8(a) shows the response times on the text-entry task (left) and the subtraction task (right); Table 5.6 contains the results of the ANOVAs. The interaction between Subtraction Difficulty and Text-Entry Difficulty was significant in both tasks. Furthermore, Subtraction Difficulty had a significant effect on the response times of the text entry

Table 5.6 ANOVA results of the pilot data. "Interaction" is the interaction between Subtraction Difficulty and Text-Entry Difficulty.

	Response Times			Accuracy		
Source	$F(1,19)$	p	η_p^2	$F(1,19)$	p	η_p^2
			Text-Entry Task			
Subtraction	69.47	< .001	.79	16.64	< .001	.47
Text-Entry	< 1	-	-	50.72	< .001	.73
Interaction	19.84	< .001	.51	17.92	< .001	.49
			Subtraction Task			
Subtraction	139.4	< .001	.88	62.98	< .001	.77
Text-Entry	32.18	< .001	.63	4.29	.052	.18
Interaction	22.73	< .001	.54	2.85	.108	.13

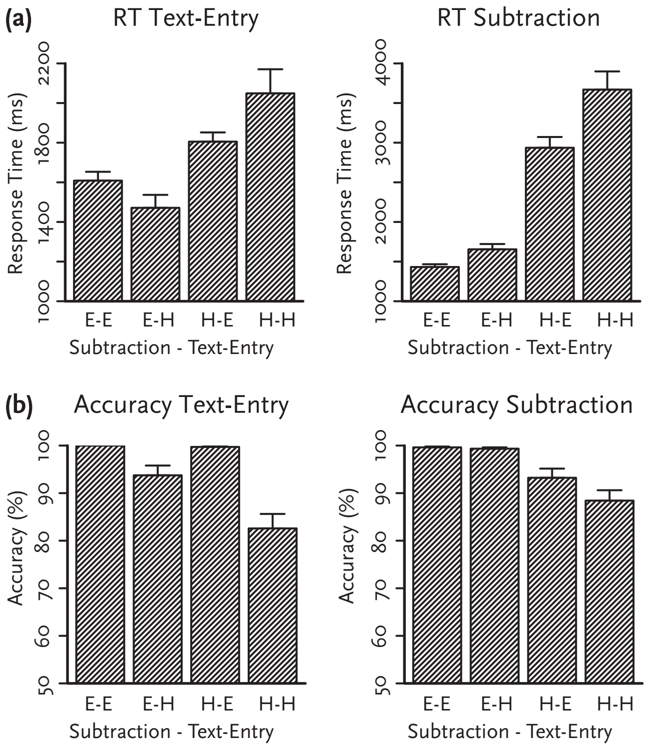


Figure 5.8 Behavioral results in the pilot experiment. RT = response time, E/E = easy–easy, E/H = easy–hard, etc., error bars indicate standard error.

task, while both Subtraction Difficulty and Text Entry Difficulty had a significant effect on the response times of the subtraction task.

Figure 5.8(b) shows the accuracies, in which the interaction effect between Subtraction Difficulty and Text-Entry Difficulty reached significance for the text entry task, but not for the subtraction task. All main effects reached significance.

An Integrated Theory of Intermediate Representations in Multitasking

In which we present our final theory of how intermediate representations are processed in the mind, and how this results in multitasking interference.

6

Chapter

Chapter 6

An Integrated Theory of Intermediate Representations in Multitasking

Abstract

In this article we propose an integrated theory of how intermediate representations – e.g., '3x = 24' when solving '3x – 6 = 18' – are processed in human multitasking. The theory, working memory in multitasking (WMM), suggests that working memory limitations are an important cause of multitasking interference. WMM states that (a) at most a single intermediate representation can be stored without decay in the so-called problem state resource, (b) representations that are not currently in the problem state resource are temporarily stored in a declarative memory store that is subject to decay, and (c) the concurrent use of multiple representations for different tasks therefore leads to interference. We will review previously published studies and present three new experiments to support the three major aspects of the theory: a single-sized problem state resource, involvement of a declarative memory store that is subject to decay, and the strategic use of intermediate representations and the environment. In addition, we will present computational models that show that the WMM theory gives a quantitative account of the observed interference effects.

Introduction

Anyone who has observed a car drifting out of its lane because the driver tries to enter a new destination in the navigation device is familiar with the negative effects that multitasking can have on performance. As early as 1931, Telford investigated interference due to human multitasking. He introduced the psychological refractory period (PRP) paradigm, and showed that people are slower to respond to the second of two tasks when these tasks have to be performed concurrently. In addition to concurrent multitasking, theorists have recently started looking at the detrimental effects that sequential multitasking (interleaving tasks) can have (e.g., Altmann & Gray, 2008; González & Mark, 2004; Monk et al., 2008; Monsell, 2003). However, both in concurrent and sequential multitasking, most studies have focused on relatively simple tasks that do not require maintaining information. In real-world tasks, which often involve multitasking, maintaining and using intermediate information is typically an important part of the task. In the current article we will therefore focus on the use of intermediate information in multitasking. We will show that the use of intermediate representations – for example $3x = 14$ when solving $3x - 5 = 9$ – is an important cause of multitasking interference, both in concurrent and sequential multitasking. To account for this kind of interference we will propose a computational theory, Working Memory in Multitasking, which yields quantitative predictions of multitasking interference due to the use, storage, and retrieval of intermediate representations.

Background

Since Telford (1931), many theories have been put forward to explain interference effects in multitasking (see for overviews, Meyer & Kieras, 1997a; Salvucci & Taatgen, 2008). Theories on multitasking can be divided into three general groups: bottleneck theories, resource theories, and cognitive control theories. Bottleneck theories assume fixed bottlenecks in human cognition that can only process one task at a time, causing interference when used by multiple tasks concurrently (e.g., Broadbent, 1958; Keele, 1973; Pashler, 1994; Welford, 1952). Theorists have identified several different bottlenecks, ranging from perceptual bottlenecks (e.g., Broadbent, 1958), to response-selection bottlenecks (e.g., Pashler, 1984; 1994), to motor bottlenecks (e.g., Keele, 1973). To unify these different bottleneck accounts, resource theories were introduced. These theories assume that attention can be flexibly employed, and that multitasking interference occurs when cognitive resources are required by multiple tasks at the same time, but not when tasks require different resources (e.g., Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1984, 2002). A third research tradition focuses on executive processing and cognitive control to explain multitasking interference (e.g., Baddeley, 1986; Cooper & Shallice, 2000; Meyer & Kieras, 1997a, 1997b; Norman & Shallice, 1986). In these theories, multitasking interference arises because of scheduling problems between tasks. That is, while tasks could in principle be carried out concurrently, executive control mechanisms enforce a certain task order, leading to interference. Using a cognitively bounded rational analysis, Howes, Lewis, and Vera (2009) have recently shown that any theory of the classical PRP effect (Schumacher et al., 1999; Telford, 1931) should contain cognitive control mechanisms, a motor bottleneck, and a response-selection bottleneck.

The recently proposed *threaded cognition* theory of multitasking indeed incorporates these elements (Salvucci & Taatgen, 2008, 2011; Salvucci, Taatgen, et al., 2009). Threaded cognition is a general theory of human multitasking that was proposed to integrate all findings to date. It assumes multiple different bottlenecks, and states that while multiple tasks can be performed concurrently, every resource in human cognition can only process one task at a time and therefore acts as a bottleneck when required by multiple tasks concurrently. Depending on the requirements of the tasks at hand, these bottlenecks lead to different patterns of interference. For instance, when two tasks need to retrieve a fact from declarative memory at the same time, threaded cognition predicts that one of the tasks will have to wait for the other task, resulting in multitasking interference. While all resources are singular in nature, the resources themselves act in parallel (cf. Byrne & Anderson, 2001). This implies that no multitasking interference will occur as long as tasks have different resource requirements (i.e. perfect time sharing, Anderson, Taatgen, et al., 2005; Hazeltine, Teague, & Ivry, 2002; Schumacher et al., 2001).

In was shown that threaded cognition can account for interference caused by two peripheral bottlenecks (vision, motor) and two cognitive bottlenecks (procedural and declarative memory; Salvucci & Taatgen, 2008; cf. Howes et al., 2009). In addition, based on its integration in the cognitive architecture ACT-R (e.g., Anderson, 2007), one more source of multitasking interference was predicted: the so-called problem state resource[1]. The problem state resource is used to store intermediate representations that are necessary for performing a task. For example, when calculating '2 + 3 × 4' mentally, one might use the intermediate representation '2 + 12'. According to the ACT-R theory, only a single intermediate representation can be maintained at a time, which should therefore lead to interference when multiple representations are required concurrently. Recently, we provided support for this prediction with a series of experiments (Borst, Taatgen, & Van Rijn, 2010). In a dual-task paradigm, subjects either needed zero, one, or two intermediate representations to perform the tasks. We showed that performance decreased considerably when subjects needed two intermediate representations at the same time, as compared to when subjects needed zero or one representations. To account for these results we developed a cognitive computational model. This model showed that both the increase in response times and the decrease in accuracy could be explained by a so-called problem state bottleneck.

Current Article

In the current article we will present an integrated theory of how intermediate representations are used in multitasking: the Working Memory in Multitasking theory (WMM). While we have previously shown that the use of multiple intermediate

[1] *The* imaginal buffer *in ACT-R terminology.*

representations at the same time leads to multitasking interference, we did not investigate how the problem state bottleneck interacts with other elements of human cognition. We will now show that the interplay of the problem state resource, a declarative memory store, and the environment can explain a much wider range of human multitasking phenomena. In the remainder of this article we will first explain the WMM theory in detail. We will then provide supporting data for three major aspects of the theory in the following three sections of the paper. Finally, we will discuss the wider implications of the WMM theory.

Working Memory in Multitasking: An Integrated Theory of Intermediate Representations in Multitasking

In this section we will describe the WMM theory,[2] which accounts for the use of intermediate representations in multitasking and the interference that can result from using them. WMM was implemented in the cognitive architecture ACT-R (e.g., Anderson, 2007; Anderson, Bothell, et al., 2004), to enable quantitative predictions of response times, errors, and neuroimaging data (see Cooper, 2007; Meyer & Kieras, 1997a; Newell, 1973; 1990 for discussions on the advantages of cognitive architectures). Using a cognitive architecture also has the advantage that interactions between central cognitive processes and perception automatically result from the modeling effort, which is crucial for modeling interactions between (often complex) tasks in multitasking (e.g., Kieras & Meyer, 1997; Van Maanen et al., 2009). We will now first discuss the main components of the WMM theory, followed by how it accounts for multitasking interference.

Main Components: The Problem State Resource and Declarative Memory

Figure 6.1 shows the main components of the WMM theory: the problem state resource and declarative memory. These elements are based on the corresponding elements in ACT-R (e.g., Anderson, 2007). The problem state resource is used to maintain intermediate representations in a task and can maintain a single intermediate representation at a time. It is assumed that a representation in the problem state resource can be used instantly, without incurring a time cost. However, it was estimated that it takes about 200 ms to store a representation in the problem state resource (e.g., Anderson, 2005). When a representation is stored in the problem state resource, its previous contents are automatically encoded in declarative memory. Representations in the problem state resource can originate from three sources: a representation can be perceived, it can be retrieved from declarative memory, or it can be the outcome of a cognitive process.

[2] *The code of models described in this article can be downloaded from http://www.jelmerborst.nl/models.*

The concept of the problem state resource stems from a series of neuroimaging experiments by Anderson and colleagues, who found that activity in the posterior parietal cortex correlates with the number of transformations of mental representations (Anderson, 2005; Anderson, Albert, et al., 2005; Anderson et al., 2003; Sohn et al., 2005). A region in the posterior parietal cortex (Montreal Neurological Institute, MNI, coordinates: -24, -67, 44; roughly indicated in Figure 6.1) is hypothesized to reflect changes to the problem state resource, while a region in the prefrontal cortex (-43, 24, 25) is supposed to reflect retrieving information from declarative memory (e.g., Anderson, 2007; Anderson et al., 2008; Borst, Taatgen, Stocco, et al., 2010).

The second major component of the WMM theory is the involvement of a declarative memory store. To this end, we used ACT-R's declarative memory store, which simulates short and long term storage of facts. In contrast to the problem state resource, it contains multiple memory items. Each item has a certain activation level, representing the strength of the item in memory. Activation of an item reflects its frequency and recency of use, and decays with a power function (Anderson & Schooler, 1991). An item that has been used more frequently in the past will have a higher activation level, as will an item that has been used more recently. The base-level activation B_i of an item at time t is calculated with the following equation:

$$B_i(t) = \ln\left(\sum_{k=1}^{n}(t-t_k)^{-d}\right) \tag{6.1}$$

in which $t_1 \ldots t_n$ indicate moments in time when the item has been (re-)created or used. The ACT-R literature has reached consensus on a value of .5 for d, the decay parameter. Thus, for each memory trace k of an item, the activation is calculated (based on how long ago k was and the decay value: $(t-t_k)^{-d}$); those activation values are summed to calculate the final activation value of an item.

Retrieving information from declarative memory is not always successful: it can either fail altogether (because activation is below the predefined retrieval threshold) or a similar but incorrect element can be retrieved (e.g., '15 – 7 = 8' when trying to retrieve a fact containing '15 – 8'). This is implemented in the ACT-R theory as a process of *partial matching*: items can be retrieved when they do not match a retrieval request completely, in which case their activation receives a mismatch penalty (see for details of partial matching, Anderson et al., 1996; Lebiere, 1999; Taatgen & van Rijn, 2011).

When retrieval of an item from declarative memory is attempted, activation of memory items is calculated with:[3]

$$A_i = B_i - MP_i + \varepsilon \tag{6.2}$$

in which B_i is the base level activation of an item calculated with Equation 1, MP_i the mismatch penalty the item receives, and ε noise drawn from a logistic distribution. Because the item with the highest activation is retrieved, noise will sometimes result in the retrieval of an incorrect (but similar) item. The probability of retrieving an incorrect

[3] *ACT-R's complete activation equation also incorporates spreading activation (see e.g., Anderson, 2007). As spreading activation does not play a role in the current article we simplified the equation.*

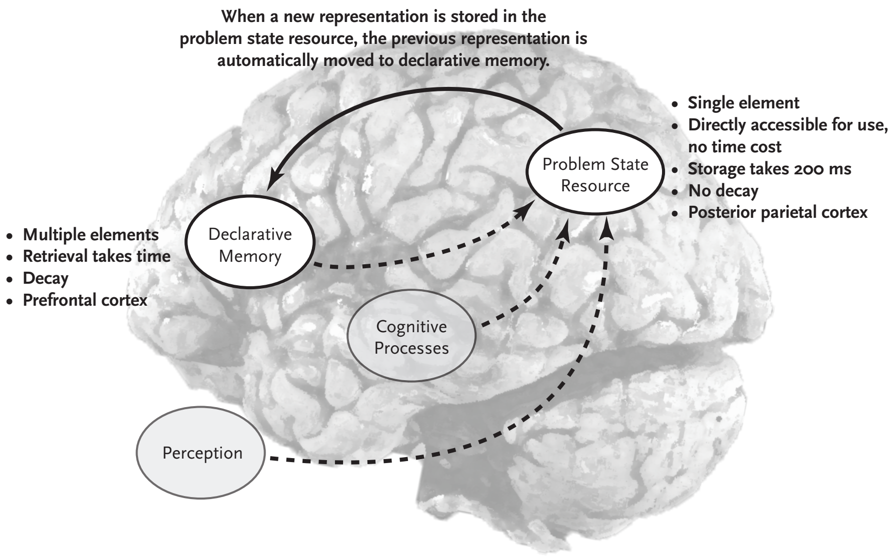


Figure 6.1 The main components of the WMM theory: the problem state resource and declarative memory. As indicated by the dashed lines, representations in the problem state resource can originate from general cognitive processes, declarative memory, or perception. Only elements and connections that we will discuss explicitly in this article are shown in the figure.

item depends on the activation difference between items: the closer the activation levels of two items, the higher the chance that an incorrect item will be retrieved.

Retrieving an item from declarative memory takes time. The duration of a retrieval depends on the activation level A_i of the item:

$$RT = Fe^{-A_i} \tag{6.3}$$

in which F is a latency scale factor (normally set between .1 and 2, Anderson et al., 1998). Thus, the higher the activation level of an item, the faster it will be retrieved from memory.

Summarizing, the main components of the WMM theory are a single-sized problem state resource and the use of a declarative memory store with memory items that are subject to decay.

Multitasking Interference

The WMM theory was developed to account for multitasking interference. Figure 6.2 illustrates the origin of multitasking interference in the theory. A dual-task situation is depicted, in which two tasks are alternated (this can be quick or slow alternation, depending on the duration of the 'General Task Actions'). Panel A shows a situation in which neither Task 1 nor Task 2 needs an intermediate representation, for example drinking coffee while casually listening to music. Only general, non-problem state related

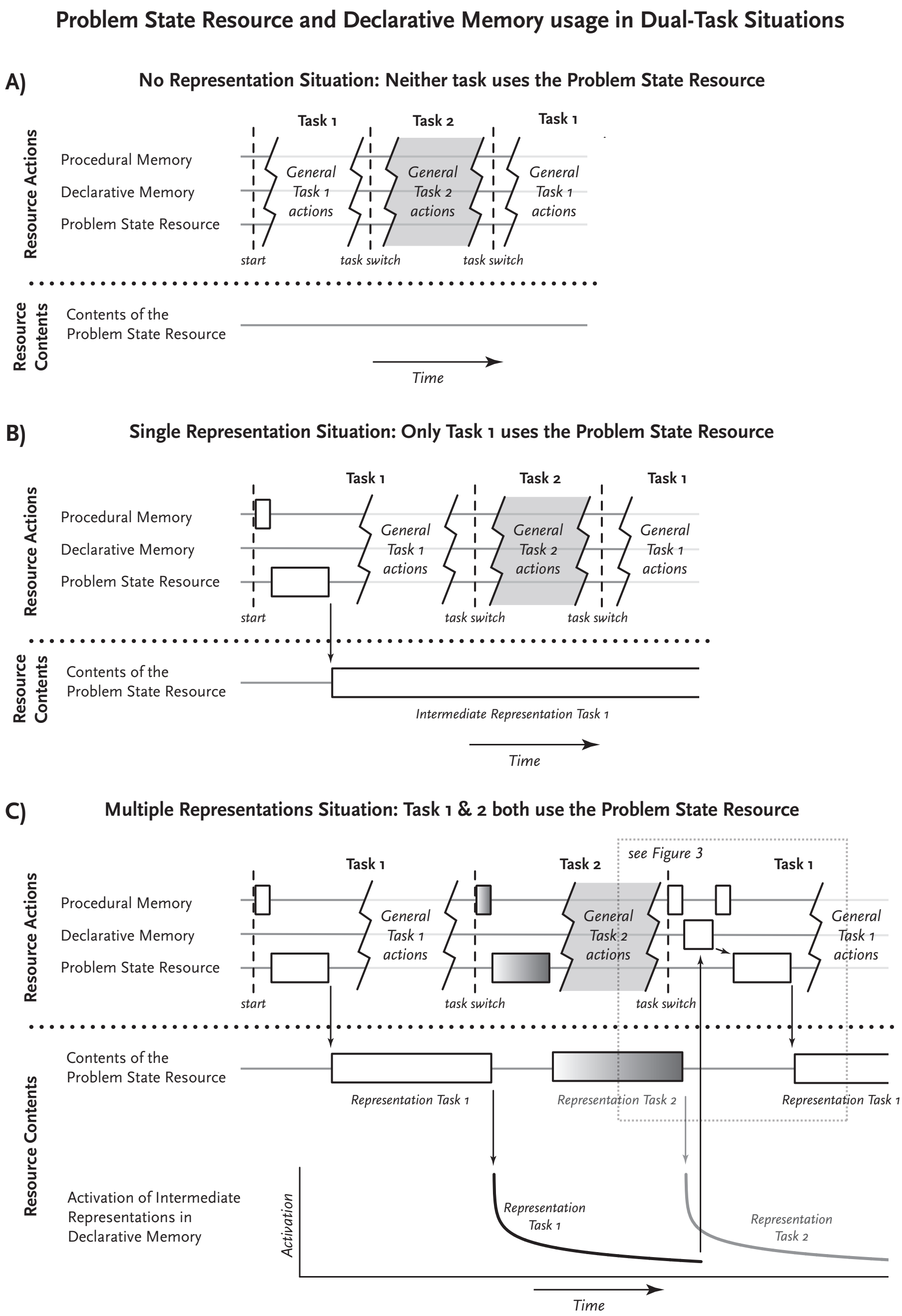


Figure 6.2 Multitasking interference in the WMM theory.

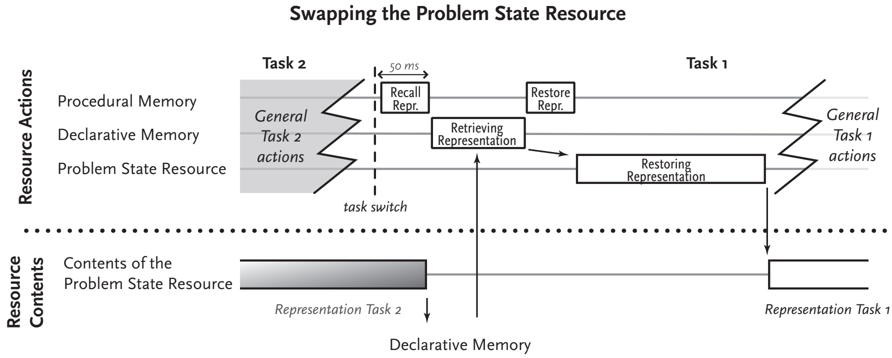


Figure 6.3 Swapping intermediate representations via declarative memory.

actions are performed for both tasks. There is thus no intermediate representation related multitasking interference in this situation.

Panel B shows a situation in which only Task 1 needs an intermediate representation. For instance, writing an article and taking a sip of coffee. At the start of Task 1, a representation – say, a reference that has to be included in the current paragraph – is stored in the problem state resource, which takes 200 ms. This representation can then be used by the 'General Task 1 actions': writing a paragraph. When the switch to Task 2 occurs, the problem state resource is not overwritten, because Task 2, drinking coffee, does not need an intermediate representation. After the next task switch, Task 1 can therefore still use its representation, and continue writing where it left off. Again, as in Panel A, there is no multitasking interference due to the use of intermediate representations.

Panel C of Figure 6.2 depicts the situation in which multitasking interference occurs: both tasks need an intermediate representation. For example, being interrupted by a phone call while writing an article. Also in this situation a representation for Task 1 is stored in the problem state resource at the start of the dual-task. However, Task 2 now also needs to use the problem state resource, and therefore stores its own intermediate representation – the topic of the call – after the switch from Task 1. By doing so, the representation of Task 1 is automatically encoded in declarative memory, where its activation starts to decay. After Task 2 finished its General Task 2 actions and the call is terminated, Task 1 recommences. However, before it can start, it first has to restore its intermediate representation from declarative memory – the reference that was relevant for the current paragraph.

This process is shown in detail in Figure 6.3. First, Task 1 notices that the problem state resource has the wrong contents, and it starts retrieving its own representation from declarative memory. This retrieval takes a certain amount of time, depending on how long ago the intermediate representation was encoded in declarative memory: the longer ago it was encoded, the longer it will take to retrieve it (calculated with Equations 6.1–6.3). When the representation is recalled, the command to restore it to the problem state resource is issued, and it is restored. Only then can Task 1 start its normal actions.

Thus, starting Task 1 after Task 2 in situation C takes 2 × 50 ms (procedural memory firing time; e.g., Anderson, 2007) + 200 ms (default value of storing an item in the problem state resource) + the retrieval time of the intermediate representation from declarative memory = at least 300 ms longer, as compared to situations A and B. The exact length of the interference depends on the duration of Task 2: the longer Task 2, the more the representation of Task 1 will have decayed in declarative memory, the longer it will take to retrieve it. This is depicted in Figure 6.4: panel A shows the decay of the intermediate representation in declarative memory while the intervening task is performed (in this case Task 2), while panel B shows the effect on the time cost of recommencing the first task. The *basic costs* represent the 300 ms of procedural memory usage and restoring the problem state resource. The other costs are influenced by the latency factor *F* (Equation 6.2), which was set to .3 for all models in this article.

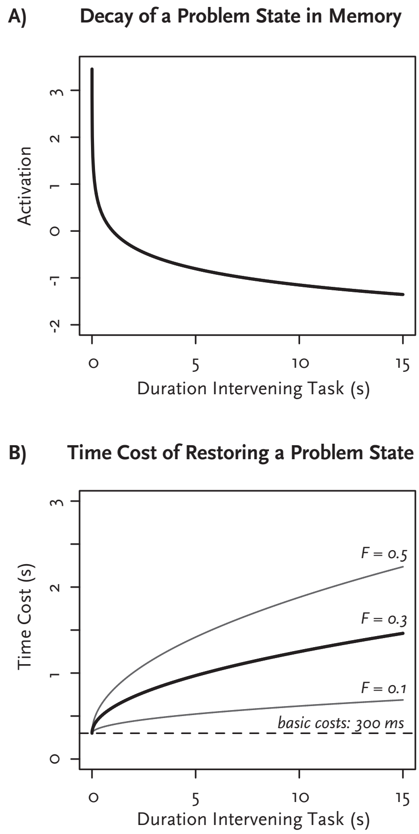


Figure 6.4 Time cost of restoring an intermediate representation to the problem state resource after an intervening task. A) shows decay of the representation in memory while the intervening task is performed, B) shows the time cost depending on the duration of the intervening task and different values of the latency factor F.

The duration of the intervening task does not only increase the time cost of restarting the first task, it also increases the probability of not being able to retrieve the intermediate representation of Task 1 at all, because its activation dropped below the retrieval threshold due to decay. This would either lead to errors because of guessing, or, when possible, to reconstructing the representation from the environment, which has its own associated costs. Even if the model retrieves a representation from memory, a less active intermediate representation will also make it more likely that a similar but incorrect representation is retrieved (see the text on partial matching and noise above).

Summarizing, the WMM theory predicts that as long as at most one intermediate representation is required no interference will occur. However, as soon as more than one representation is needed, it predicts interference both in response times and errors due to a single-sized problem state resource. Moreover, the interference costs increase with the duration of the intervening task, because representations in declarative memory are subject to decay.

Strategic use of the Problem State Resource and the Environment

The WMM theory explains how mentally maintaining intermediate representations can lead to multitasking interference. However, Figure 6.1 shows that an intermediate representation can also originate in perception, opening up the possibility to use the environment as an external memory, and thereby avoiding the limits of the problem state resource. For example, when solving a multicolumn subtraction problem on paper it is not uncommon to indicate whether a carry is in progress, which decreases the problem state resource requirements. Even when one relies on memory, and stores the carry in the problem state resource, there are two possible strategies to continue after losing the representation due to an interruption: recalling whether a carry was in progress from declarative memory, or reconstructing it by recalculating the previous column. While reconstruction is the safer option, it is also likely to take more time than recall from memory.

We assume that these are strategic processes: when reconstructing or perceiving a representation from the environment takes less time than a cognitive 'in-the-head' strategy (using the problem state resource as it is, or retrieving it from memory), humans will use the environment, otherwise they will use a cognitive strategy. This is in accordance with the Soft Constraints Hypothesis, which proposes that our strategy choices aim at minimizing temporal costs instead of, for example, mental effort (Gray & Fu, 2004; Gray, Sims, Fu, & Schoelles, 2006). To support this, Gray et al. showed in four experiments that subjects always used the fastest local strategy, instead of minimizing mental effort or total time-on-task. Even when minimizing local time led to suboptimal behavior (using fast but imperfect knowledge in-the-head vs. slower but perfect knowledge in-the-world) or to more memory effort (memorizing multiple facts instead of perceptually revisiting a display), subjects opted for the fastest method (that is, the fastest *local* method, as this often relied on imperfect knowledge in-the-head it could lead to mistakes and longer total time-on-task). The assumption of the WMM theory is that the Soft Constraints Hypothesis also holds for processing intermediate representations, and therefore that people always choose the fastest way of performing a task, whether that is using the problem state resource, declarative memory, or the environment.

Overview of the Article

In the remainder of this article we will provide support for the following three aspects of the WMM theory:

- the single-sized problem state resource, leading to multitasking interference when more than one representation is required for a task;

- the involvement of a declarative memory store with decay, leading to more interference with longer task interruptions;
- the strategic nature of using the problem state resource, declarative memory, and the environment.

We will discuss support for these aspects of the theory in the next three sections of this article.

The Single Element Size of the Problem State Resource

In this section we will review an experiment that supports the assumption that the problem state resource can contain at most one representation at a time. The experiment shows that using the problem state resource for two tasks at the same time leads to interference. In addition to the behavioral data we will also present model fits, to show that the WMM theory can account for the patterns in the behavioral data. The experiment has been published before as Experiment 1 in Borst, Taatgen, and Van Rijn (2010).

To test whether the problem state resource can at most contain a single intermediate representation, a dual-task experiment was designed in which subjects had to alternate between solving multi-column subtraction problems and entering text. Both tasks had two versions: an easy version in which no intermediate representations were required and a hard version in which subjects needed intermediate representations to perform the task. The experiment had a 2 × 2 factorial within-subjects design: Subtraction Difficulty (easy/hard) × Text-Entry Difficulty (easy/hard).

Experimental Design

Figure 6.5 shows a screenshot of the experiment. On the left side of the screen subjects had to solve 10-column subtraction problems in standard right-to-left order. In the easy version of the subtraction task the lower term was always smaller or equal to the upper term in each column, meaning that subjects never had to carry between columns. In the hard version (depicted in Figure 6.5) subjects had to carry in 6 out of 10 columns. The assumption is that subjects used their problem state resource to keep track of whether a carry was in progress.

On the right side of the screen subjects had to enter 10 letter strings using the mouse. In the easy, no problem state version these strings were presented letter by letter (an 'I' is presented in Figure 6.5). Subjects had to click on the corresponding button, after which the next letter appeared. In the hard version subjects had to enter 10-letter words without feedback. That is, at the start of a trial a complete 10-letter word was presented, but as soon as they clicked on the first letter, the word disappeared (i.e. subjects could neither see what word they were entering, nor what they had already entered). It was assumed that subjects used their problem state resource to maintain the word and the position in the word (e.g., 'information, 4th letter').

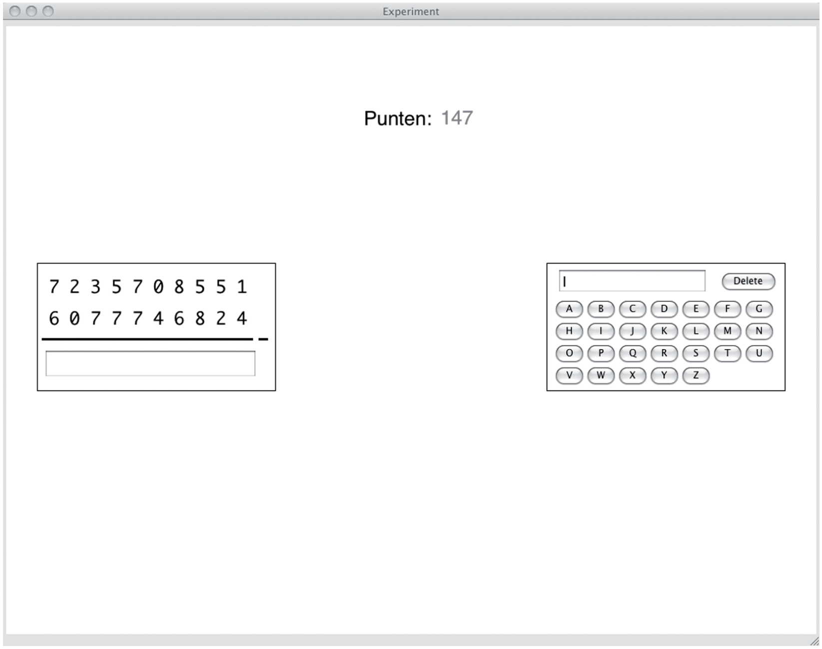


Figure 6.5 Screenshot of Experiment 1. Adapted with permission from Borst, Taatgen, and Van Rijn (2010). Copyright 2010 by the American Psychological Association.

The interesting part of the experiment is that subjects had to alternate between the two tasks after every digit and letter. Thus, they had to maintain the intermediate representations – the carries and words – while giving a response on the other task. According to the WMM theory, this should result in interference in the *hard subtraction – hard text-entry* condition, because then the problem state resource has to be swapped out on each step in a trial as it can only contain a single element (situation C in Figure 6.2). In all other conditions at most one intermediate representation is required to do the tasks, which, according to the theory, does not result in interference (*easy – easy*: Figure 6.2A, *easy – hard* and *hard – easy*: Figure 6.2B). See Borst, Taatgen, and Van Rijn (2010) for further details of the methods.

Results

Figure 6.6 shows the results of the experiment (black bars). The top two panels show the response times, the bottom panels accuracy. As predicted by the WMM theory, both in the text-entry task (left) and in the subtraction task (right) there is a clear increase in response times in the *hard-hard* condition as compared to all other conditions. In the response times of the subtraction task there is also a clear effect of subtraction difficulty: response times increase when subtraction becomes hard. However, in the *hard-hard* condition response times are even higher. These results were confirmed by

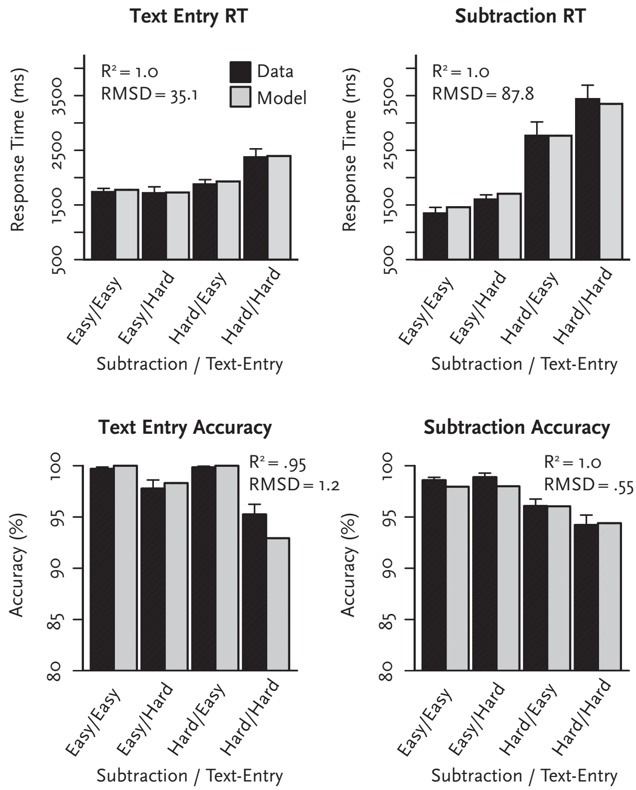


Figure 6.6 Results of Experiment 1. Error bars represent standard errors. RMSD = root-mean-square deviation; RT = response time. Adapted with permission from Borst, Taatgen, and Van Rijn (2010). Copyright 2010 by the American Psychological Association.

significant interaction effects between Subtraction Difficulty and Text-Entry Difficulty for both tasks.

In the accuracy data (Figure 6.6, lower panels, black bars) similar effects were observed. Accuracy decreased with task difficulty of the tasks itself, but even more in the *hard-hard* conditions. For instance, in the subtraction task subjects hardly made any mistakes as long as subtraction was easy. When subtraction was hard the amount of mistakes increased, but it increased even more when text-entry was hard as well. The same effects can be seen in the text-entry data. Statistically, the interaction effect between Subtraction Difficulty and Text-Entry Difficulty was significant for the subtraction task, and it showed a trend towards significance for the text-entry task.

Model

To see whether a single-sized problem state resource could have caused the effects in the data, we implemented a computational model of the task. The grey bars in Figure

6.6 show the results. Based on general characteristics of ACT-R, response times and accuracy data in the *easy-easy*, *hard-easy*, and *easy-hard* conditions were estimated (see for details, Borst, Taatgen, & Van Rijn, 2010). The WMM theory adds the interference effects in the *hard-hard* condition: increased response times and decreased accuracy. The increased response times are accounted for by swapping the contents of the problem state resource on each step of a trial, as explained above (Figure 6.2C). The model produces the decrease in accuracy by sometimes retrieving an older, incorrect representation from memory when swapping the problem state resource.

Discussion

The experiment was performed to test whether the problem state resource can contain a single or multiple intermediate representations. As predicted, the data show that when more than one representation is needed concurrently, performance decreases considerably. The model fit shows that the WMM theory can account both for the effects on response times and on accuracy.

One thing to note is that, instead of using a mental strategy, it is possible to reconstruct the intermediate representation of the subtraction task by reprocessing the previous column (see Figure 6.5). If subjects had used this strategy in all conditions, we would not have found a difference between the *hard subtraction – easy text-entry* and the *hard subtraction – hard text-entry* conditions. However, it is conceivable that subjects reconstructed the intermediate representation in the *hard-hard* condition, but not in the *hard-easy* condition. This would have resulted in the same effects on response times: response times would have been higher in the *hard-hard* condition because of the costs of reconstructing the representation. We controlled for this in Experiment 3 in Borst, Taatgen, and Van Rijn (2010) and in several other experiments (e.g., Borst, Taatgen, Stocco, et al., 2010) by masking previous columns. These experiments yielded comparable results as the current experiment. This is in accordance with the prediction of the WMM theory that reconstruction via the environment only takes place when it takes less time than recalling a representation from memory, as the model predicts that reconstruction would take longer than the observed interference effect of 400 ms (see for more on the strategic use of the problem state resource the section 'The Strategic Nature of using the Problem State Resource and the Environment' below). Furthermore, reconstruction would not have explained the effects in the accuracy data: there is no reason why accuracy would be lower in the *hard-hard* condition in that case.

The presence of interference in the *hard-hard* condition indicates that the difficulty of one task affects both tasks. The simplest explanation for this interference is a resource that is shared by both tasks. As the tasks are never performed concurrently, this is probably a resource that is used while doing the other task. The obvious candidate for that is the resource that maintains intermediate information required by the tasks, the problem state resource in the WMM theory. According to the theory, the interference effects in the data are caused by swapping out a single-sized problem state resource. However, there are at least two alternative explanations to explain the data of this experiment: memory load (e.g., Logan, 1979; Woodman et al., 2001) and a phonological loop bottleneck. To test these possibilities, Borst, Taatgen, and Van Rijn (2010) conducted two additional experiments. These experiment clearly showed that these alternatives did not hold, and they concluded that a single-sized problem state resource was the most likely explanation of the results.

Summary

In this section we set out to support the assumption of the WMM theory that the problem state resource can only maintain a single representation at a time. The experiment that we reviewed showed that when subjects had to maintain more than one representation at a time, considerable interference occurred. These results are in line with the WMM theory, and imply a single-sized problem state resource.

The WMM model showed a nice fit to the results: it predicted interference effects both in response times and accuracy data, which were indeed observed. Fitting the model to the data showed that it could also account for the size of the effects. This is not surprising, given that there are two parameters that determine the size of the time cost. Figure 6.4 shows that this time cost is determined by the basic costs and the declarative memory latency factor *F*, which scales the time of retrieving a representation from declarative memory. To fit the model to the data we set the latency factor *F* to .3 for both experiments.

In the next section we will look at the assumption of the WMM theory that there are two separate memory stores involved in processing intermediate representations: the problem state resource and a declarative memory store.

Two Memory Stores for Intermediate Representations

In the previous section, we reviewed an experiment that supports a single-sized problem state resource. According to the WMM theory, this single-sized problem state resource is a separate resource from declarative memory. Where declarative memory is assumed to consist of multiple elements that take time to retrieve and are subject to decay, the problem state resource can only contain a single element, which is directly accessible and not susceptible to decay. In this section we will discuss support for this assumption. We will first review fMRI data that shows that problem state activity is best represented in a different brain area than declarative memory activity, supporting the view of two separate memory stores. We will then turn to two new experiments, that show that (1) the WMM theory explains interference effects of interruptions, and (2) a declarative memory resource with decay is needed to account for these effects, in addition to a single-sized problem state resource.

Neuroimaging Evidence for Two Separate Memory Stores

To investigate the neural correlates of the WMM theory we conducted two fMRI experiments. Based on ACT-R's predefined resource-brain mapping (e.g., Anderson, 2007; Anderson et al., 2008), we first conducted confirmatory region-of-interest analyses. These analyses showed that the WMM theory could make plausible *a priori* fMRI predictions (Borst, Taatgen, Stocco, et al., 2010; Borst, Taatgen, Van Rijn, Stocco, & Fincham, 2009). These predictions were based on the assumption that declarative memory and the problem state resource are separate entities, and that they are reflected by activity in different brain areas (e.g, Anderson et al., 2007). To test this assumption, we subsequently applied a novel exploratory fMRI analysis method to the data (Borst, Taatgen, & Van Rijn, 2011). This method, termed *model-based fMRI*, confirmed that the problem state resource and declarative memory are best represented by two different brain areas (posterior parietal and prefrontal cortex, respectively). We will now discuss this experiment and analysis in more detail.

Design

To investigate the neural correlates of the WMM theory, we conducted the subtraction – text-entry experiment in an fMRI scanner. That is, the interface was adapted to minimize eye- and head-movements, and all responses now had to be given with a mouse, but the basic task remained unchanged. Thus, subjects had to constantly alternate between solving 10-column subtraction problems (of which only one column was shown at a time in this experiment) and entering 10-letter strings. Both tasks again had an easy, no intermediate representation condition, and a hard condition with intermediate representations (see for details, Borst, Taatgen, Stocco, et al., 2010).

Analysis

The behavioral data showed similar effects as the data reviewed above, further corroborating the single-sized problem state resource hypothesis (Borst et al., 2011). To analyze the fMRI data, we applied a novel fMRI analysis technique, which has previously been used to investigate the neural mechanisms of reinforcement learning (e.g., Gläscher & O'Doherty, 2010; O'Doherty et al., 2007). Instead of regressing the conditions of the experiment against the experimental data, as is traditionally done in fMRI research (e.g., Friston et al., 2007), we regressed the activity of the model's resources directly against the fMRI data.

To that end we first fitted the WMM model to the behavioral data: We performed model simulations of each individual subject, using the same stimuli, in the same order, as the subject received. In addition, we lined up the key-presses of the model to the key-presses in the data, resulting in a perfect model-behavioral data fit. We then convolved the activity of the model's resources with a hemodynamic response function (describing the sluggish brain response that is measured with an fMRI scanner; e.g., Friston et al., 2007), and entered the resulting signal into General Linear Models (GLM). This results in regressing the model's resource activity directly against the brain activity in all voxels of the brain, and shows which voxels correlate significantly with the predicted activity of the model's resources. Thus, it shows where in the brain the resources of a model are most likely represented.

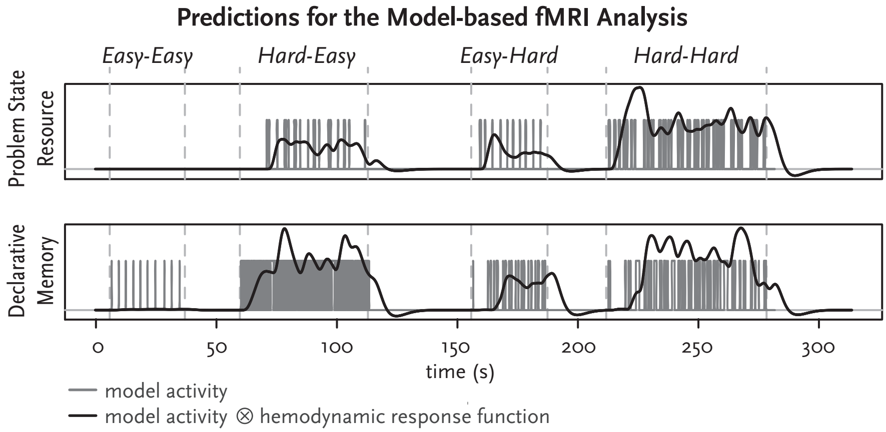


Figure 6.7 Model activity and associated hemodynamic response functions that were used for the fMRI analysis. Note that declarative memory usage in the easy-easy condition is so fast that it hardly causes a predicted hemodynamic response.

Results

Figure 6.7 shows the model activity and the predicted hemodynamic response for the problem state resource and declarative memory over the course of four trials (note that this is just an example, the exact signal was different for each trial depending on the stimuli in a trial; a trial constitutes solving a complete subtraction problem and entering a 10-letter string). The grey line shows the activity of the resources. The problem state resource is not used in the *easy-easy* condition, used for one of the tasks in the *easy-hard* and *hard-easy* conditions, and used most in the *hard-hard* condition: not only is it used for both tasks, but its contents are also swapped on each step of a trial. Declarative memory shows a very similar pattern. First, declarative memory is hardly used in the *easy-easy* condition (as the subtraction task only requires simple, and thus fast, retrievals, e.g., '5 – 2 = 3'). Second, we see increased levels in the *hard subtraction – easy text-entry* condition, because in that condition additional retrievals are necessary for processing the carries (e.g., '5 – 8 = -3', but also '15 – 8 = 7'). Third, declarative memory is used for the spelling processes of the words in the text-entry task (Borst, Taatgen, Stocco, et al., 2010; Borst, Taatgen, & Van Rijn, 2010), resulting in increased levels in the *easy subtraction – hard text-entry* condition. Finally, it is used most in the *hard-hard* condition, on the one hand for the tasks themselves, and on the other hand for retrieving a representation to reconstruct the contents of the problem state resource on each step of a trial. The black line in Figure 6.7 shows the convolution of the resource activity with the hemodynamic response function, yielding a very detailed prediction of brain activity (such a prediction was made for all subjects over all trials in the experiment). This signal was subsequently regressed against the brain data: showing which areas of the brain correlate significantly with these predicted signals. As the predictions for the problem state resource and declarative memory were very similar, it was relatively unlikely to find different brain areas for these resources.

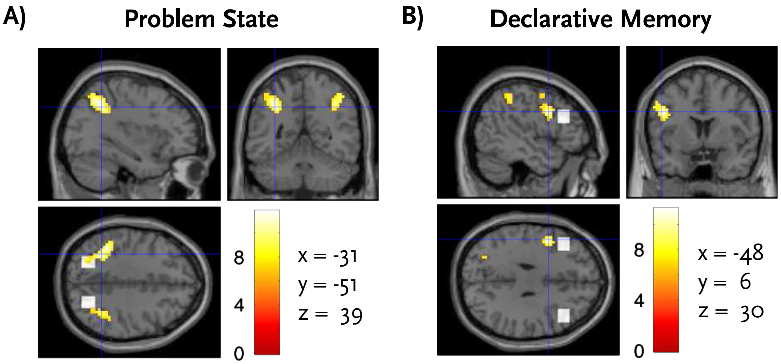


Figure 6.8 Results of the fMRI analysis: the best fitting areas for a) the problem state resource and b) declarative memory. The white squares indicate ACT-R's predefined regions (e.g., Anderson, Fincham, Qin, & Stocco, 2008), xyz-coordinates are of the most significant voxel.

Figure 6.8 shows the results. The best fitting area for the problem state resource was located in the inferior parietal lobule, around the intraparietal sulcus (Borst et al., 2011). This area has not only been linked to ACT-R's problem state resource in the past (e.g., Anderson, 2005; Anderson, Albert, et al., 2005; Anderson et al., 2003; Sohn et al., 2005), but has also been found in other studies on working memory (e.g., LaBar et al., 1999; Smith et al., 1998; Wager & Smith, 2003). Declarative memory correlated best with a region in the prefrontal cortex, around the inferior frontal gyrus. This area is known to be involved in retrievals from memory (e.g., Cabeza et al., 2002; Fletcher & Henson, 2001; Wagner et al., 2001). Furthermore, when we lowered the significance threshold, it became apparent that both areas are part of a larger fronto-parietal network, a network that is often implicated in working memory research (e.g., Collette et al., 2006; Collette & Van der Linden, 2002).

Discussion

While the predicted hemodynamic response functions of the problem state resource and declarative memory were quite similar, the current analysis shows that they correlate best with different areas in the brain (see for a detailed discussion of the power of model-based fMRI, Borst et al., 2011). That we found two significant, spatially quite different areas, suggests that the functions of those resources are indeed best represented by two separate memory stores as the WMM theory assumes. In this scheme, the problem state resource is used for maintaining and updating a single intermediate representation for the task at hand, while declarative memory is used for storage of information when it is not directly available.

If it is indeed the case that intermediate representations are swapped in and out of the problem state resource via declarative memory, we should find a typical declarative memory effect on these representations in memory: decay. In the next section we will discuss two experiments that test this prediction.

Interruption Experiments: Decay in Declarative Memory

One major assumption of the WMM theory is that intermediate representations are swapped in and out of the problem state resource via declarative memory. Because the activation of a representation in declarative memory decays while an intervening task is performed (Figure 6.2C), this leads to the prediction that the longer an intervening task lasts, the higher the costs of restoring an intermediate representation will be (see Figure 6.4). To test this prediction, we conducted two 'interruption experiments' (e.g., Gillie & Broadbent, 1989).

It is well known that interruptions of a task lead to a decrease in performance (e.g., Gillie & Broadbent, 1989; McFarlane & Latorella, 2002). Two important factors that determine the disruptiveness of interruptions are the duration and the complexity of the interrupting task. The longer the interrupting task, the longer it takes to resume the primary task (Hodgetts & Jones, 2006; Monk et al., 2008), and the more complex the interrupting task, the longer the resumption time (e.g., Gillie & Broadbent, 1989; Hodgetts & Jones, 2006; Monk et al., 2008; Zijlstra, Roe, Leonora, & Krediet, 1999). This maps well onto the WMM theory, if we assume that an important factor in the disruptiveness of interruptions is the loss of intermediate representations of the primary task and subsequent decay of these representations in memory. According to the WMM theory, intermediate representations of the primary task would only be disturbed when the *interrupting* task is sufficiently complex to need an intermediate representation for itself, at least partly explaining the complexity effect. At the same time, the longer the interrupting task, the further a representation would have decayed in declarative memory, leading to increased resumption times. The latter idea is similar to the memory for goals model (Altmann & Trafton, 2002; Salvucci, Monk, et al., 2009; Trafton et al., 2003).

The hypothesis that intermediate representations are an important factor in the disruptiveness of interruptions leads to an interesting new prediction: the duration of an interruption should only influence the resumption time of the primary task *if both tasks need an intermediate representation*. If the primary task does not need an intermediate representation, there should be no resumption costs other than costs for re-attending the task, and these costs should not increase with the duration of the interruption. If the secondary task does not need an intermediate representation, the same holds: an intermediate representation of the primary task can now be maintained throughout the interruption, enabling the primary task to be continued directly after the interruption (cf. Figure 6.2B). Only when both tasks need an intermediate representation (Figure 6.2C), there should be an increase of resumption costs with the length of the interruption, in addition to the extra costs of restoring the representation to the problem state resource itself (basic costs in Figure 6.4).

To test this prediction we conducted two interruption experiments. In these experiments we interrupted a primary task with a secondary task. As before, both tasks had two conditions: and easy condition that did not require an intermediate representation, and a hard condition that did require the use of an intermediate representation. In addition, we now varied the length of the interruptions, to test

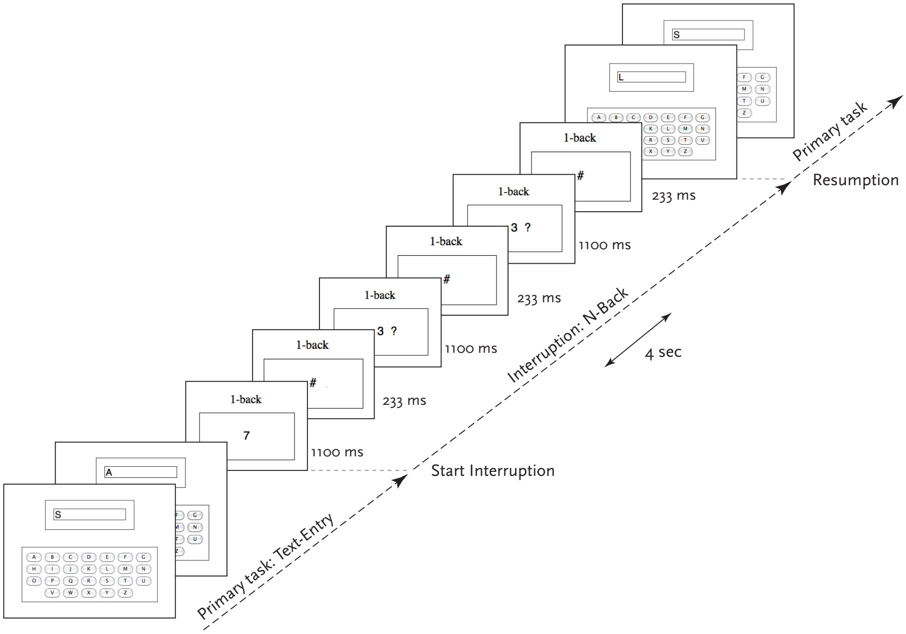


Figure 6.9 Setup of Interruption Experiment 1, in which the text-entry task is interrupted by an n back task. The figure shows the easy text-entry – easy n-back condition, with a 4 second interruption.

whether the costs of restoring a representation increase with the duration of an interruption. If that were the case, it would imply that intermediate representations are indeed swapped in and out of the problem state resource via a declarative memory store that is subject to decay.

Interruption Experiment 1: Text-Entry and N-Back

In the first interruption experiment, we used the text-entry task from the previous experiments as the primary task, and interrupted it twice every trial with a so-called n-back task. Figure 6.9 shows the setup of the experiment: subjects started with the text-entry task, which was unpredictably interrupted by the n-back task. After doing a varying number of steps in the n-back task, the text-entry task recommenced.

The text-entry task was the same task as described earlier: in the easy condition subjects were presented with a letter, they had to click on the corresponding button, followed by the next letter, etc. In the hard version, subjects had to enter a 10-letter word without feedback.

In the n-back task (Kirchner, 1958), a rapid stream of digits was sequentially presented to the subjects (Figure 6.9). Each number was on the screen for 1100 ms, followed by a mask (a #-mark) for 233 ms. In the easy, no representation condition, subjects had to do a 1-back task: they had to indicate whether the current number was the same or different as the previous number. A response had to be made while the stimulus was

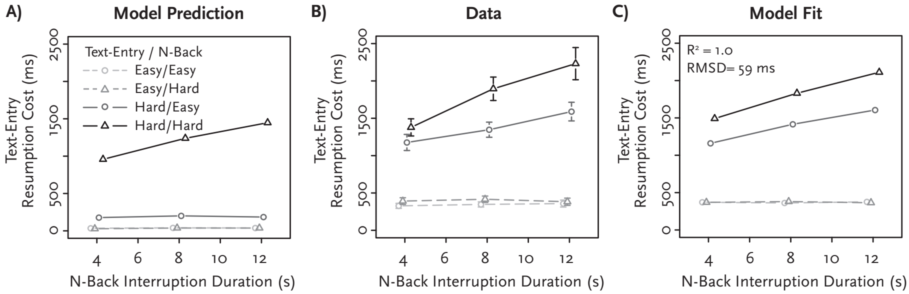


Figure 6.10 Model prediction (A), data (B), and model fit (C), of the resumption costs in Interruption Experiment 1.

shown, within 1100 ms. No response was required to the first number. As the presented mask was very short, this did not require the use of an intermediate representation: subjects could simply judge whether the shape was the same or different before and after the short mask. In the hard version of the task, subjects had to do a 2-back task: they had to judge whether the current number was the same as two numbers back. In this condition no response was required on the first two steps in the n-back task. Now subjects had to use intermediate representations to perform to the task: they constantly had to keep track of what the number two back was. When subjects made a mistake on the n-back task, a loud buzzer sounded, reminding the subject to focus on the task.

Each text-entry trial (entering a 10-letter string) was interrupted twice by the n-back task. The points of interruption were varied between the 2nd letter and the 9th letter, that is, subjects always started, and also ended, with typing at least two letters in the text-entry task. There were also at least two letters between the two interruptions. This gave 15 different trial-types: the points of interruption were therefore unpredictable to the subjects. The length of the interruptions was 3, 6, or 9 n-back steps, thus 4, 8, or 12 seconds. There was no relation between the length of the first and the second interruption in a trial. The experiment had a 2 × 2 × 3 factorial within-subjects design (Text-Entry Difficulty (easy/hard) × N-Back Difficulty (easy/hard) × Interruption Duration (4/8/12 sec). Additional methods are reported in Appendix A.

Predictions. Figure 6.10A shows predictions of a WMM model for this task. Resumption cost is plotted against interruption duration, which is the duration of the n-back task. Resumption cost is the extra cost after an interruption as compared to normal responses. We calculated resumption costs by subtracting the average response time of responses that did not follow an interruption in a condition from the response time of responses immediately following an interruption.

According to the WMM theory – and assuming that intermediate representations are the only factor in the disruptiveness of interruptions – as long as the text-entry task is easy (the dashed lines in Figure 6.10) and does not need an intermediate representation, there is no effect of the interruptions, independent of the difficulty of the n-back task. When text-entry is hard and the n-back task is easy, a small resumption cost is predicted, which does not increase with interruption duration. This cost is

Table 6.1 ANOVA results for resumption cost in Interruption Experiment 1.

Source	$F(1,15)$	p	η_p^2
Text-Entry Difficulty	155.96	< .001	.91
N-Back Difficulty	18.92	< .001	.56
Interruption Duration	33.11	< .001	.69
Text-Entry × N-Back	16.24	.001	.52
Text-Entry × Interruption Duration	22.04	< .001	.60
N-Back × Interruption Duration	3.75	.07	.20
Text-Entry × N-Back × Interruption	3.18	.09	.17

not related to problem state resource updates, but originates from the model already knowing which letter it has to enter next in the hard text-entry condition (in contrast to the easy text-entry condition, in which the next letter has to be perceived from the screen). It can therefore start preparing the next response while still executing the motor action of the current response. However, this is not possible during an interruption, leading to slightly longer response times directly after an interruption. In the *hard-hard* condition a large cost is predicted that increases with interruption duration. This is due to the single-sized problem state resource: because the n-back task also needs intermediate representations when it is hard, the intermediate representation of the text-entry task has to be restored after the interruption. This cost increases with interruption duration, as the representation will have decayed further in declarative memory the longer the interruption lasts, and it will therefore take more time to retrieve it again (see Figure 6.4).

Results. Accuracy on the n-back task was in all conditions over 90%, indicating that subjects focused on the n-back task. Figure 6.10B shows the resumption costs of the text-entry task. The first thing to note is that as long as text-entry is easy, there are no increasing costs with interruption duration. However, there are resumption costs of about 400 ms in all conditions, which the model did not predict. When text-entry was hard, resumption costs were much higher. Where the model predicted a small cost for *hard text-entry – easy n-back*, the data show a much larger effect than predicted, which furthermore increases with interruption duration. When both tasks were hard there is an additional increase in resumption costs, which increases slightly steeper with interruption duration than when n-back was easy. The ANOVA largely confirmed these results: besides main effects of Text-Entry Difficulty, N-Back Difficulty, and Interruption Duration, also the two-way interaction effect between Text-Entry Difficulty and N-Back Difficulty was significant. The three-way interaction between Text-Entry Difficulty, N-Back Difficulty, and Interruption Duration showed a trend towards significance. The ANOVA results are reported in full in Table 6.1; detailed analysis procedures are reported in Appendix A.

Discussion. Interruption Experiment 1 was conducted to test the assumption of the WMM theory that intermediate representations are processed via a declarative memory store that is subject to decay. The WMM theory predicted that longer interruptions of a task lead to higher resumption costs, but only when both the primary and the interrupting task require an intermediate representation. Indeed, the experiment showed that as long as the primary task did not need an intermediate representation, there were costs due to the interruption, but these costs did not increase with the interruption duration. While this was in accordance with our predictions, it runs counter to one of the standard effects described in the literature: increasing cost with interruption length (e.g., Hodgetts & Jones, 2006; Monk et al., 2008). When the primary task needed an intermediate representation, we found much higher resumption costs, which increased with interruption duration. In addition, when both the primary and the secondary task required an intermediate representation, these costs were even higher. According to the WMM theory, this is due to the single-sized problem state resource, necessitating the restoration of the representation after an interruption in the *hard-hard* condition. As these costs increase with interruption duration, this argues in favor of the WMM theory's assumption that intermediate representations are swapped in and out of the problem state resource via a declarative memory store with decay.

However, while the WMM theory predicted a low cost and a flat line in the *hard text-entry – easy n-back* condition, the data show a high resumption cost with a clear increase with interruption duration. There are at least three possible explanations for this discrepancy between the predictions and the data. One possibility is that the easy n-back task required the use of an intermediate representation, leading to interference in the *hard text-entry – easy n-back* condition. However, in that case we would not expect a difference between the results of the *hard text-entry – easy n-back* and the *hard text-entry – hard n-back* conditions, as they both would be effectively *hard-hard*. A second possibility is that the problem state resource sometimes looses the stored representation. For instance, Anderson and Qin (2008) proposed that representations were lost on average every 20 seconds. While such a time-scale does not work for the current model, a similar mechanism might explain the data: a representation would sometimes (but not always) be lost during an interruption in the *hard text-entry – easy n-back* condition, leading to resumption costs that lie between the model's prediction and the costs in the *hard-hard* condition. A third possibility is that the text-entry task requires two intermediate representations instead of one: one for the word ('information') and one for the position in the word ('4th letter'). This would mean that one of these representations decays in the *hard text-entry – easy n-back* condition (as the problem state resource can only be used to maintain one of the representations), leading to increased costs with interruption duration, while both representations decay in the *hard-hard* condition, leading to even higher costs.

To see whether this last possibility can explain the data we implemented it as a WMM model. When the model is interrupted in the *hard text-entry – easy n-back* condition it stores the word in the problem state resource, while the position in the word decays in declarative memory. In the *hard-hard* condition, neither the word nor the position in the word can be stored in the problem state resource, and therefore they both decay in declarative memory. Figure 6.10C shows the results: the model fits the patterns in the data, making a dual-representation strategy a possible explanation of the data.[4] For this model fit we implemented a dual-representation strategy for the text-entry task and estimated the general costs of interruptions from the data, which were added to the outcome of the model. Thus, we took the average of the resumption costs in the easy text-entry conditions, and added this to the model results. This is not meant to give an explanation of these costs, but only to ensure that we estimated the costs due to intermediate representations correctly.

While a dual-representation strategy seems to be a possible way of explaining the data, we cannot distinguish between this explanation and the other possible explanations discussed above. To test whether it is a more likely explanation than the other explanations, we performed a second interruption experiment, in which we made sure that only a single representation was needed for the primary task.

Interruption Experiment 2: Subtraction and N-Back.

In the second interruption experiment we tested whether a dual-representation strategy is a probable explanation for the data of Interruption Experiment 1. Instead of text-entry, we now used the subtraction task described earlier as the primary task, again interrupted with the n-back task. If the dual-representation strategy for the text-entry task explains the data above, then we should find a flat line in the *hard subtraction – easy n-back* condition (cf. Figure 6.10A), as only a single representation is necessary for the subtraction task. If one of the other explanations is more likely, we should find a pattern similar to the one that was found for the text-entry task.

The only difference between the two experiments is the use of the subtraction task as the primary task instead of text-entry. A complete 10-column subtraction problem was shown on the screen, but solved columns were masked with #-marks. Interruption Experiment 2 thus had a 2 × 2 × 3 factorial within-subjects design (Subtraction Difficulty (easy/hard) × N-Back Difficulty (easy/hard) × Interruption Duration (4/8/12 sec). Additional methods are reported in Appendix A.

Results. Accuracy on the n-back task was above 85% in all conditions. Figure 6.11A shows the resumption costs in the different conditions; Figure 6.11B the model fit. Again, no effect of interruption duration on the resumption costs was found as long as subtraction was easy. In addition, when subtraction was hard and n-back easy, we also did not observe increasing costs with interruption duration, unlike in the previous experiment. Only when both tasks required an intermediate representation, in the *hard-hard* condition, we observed increasing costs with interruption duration. The ANOVA (Table 6.2) confirmed main effects of Subtraction Difficulty and N-Back Difficulty. Furthermore, also the interaction effect between Subtraction Difficulty

[4] *If a dual-representation strategy is indeed the correct way of describing behavior in the text-entry task, the question is whether the model fits of Borst, Taatgen, and Van Rijn (2010) and the model-based fMRI results of Borst et al. (2011) still hold. We implemented the dual-representation strategy also for these experiments, which yielded similar results as before. We report these analyses in Appendix B.*

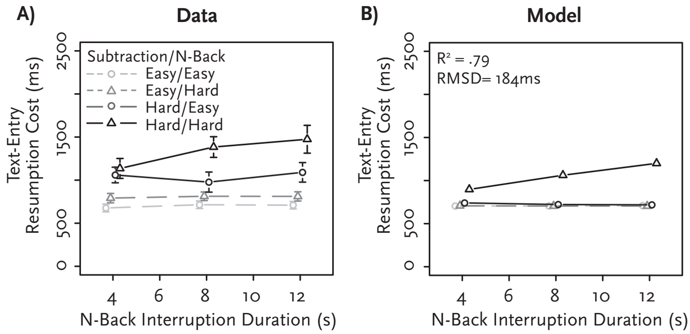


Figure 6.11 Resumption costs (A), and model fit (B) of Interruption Experiment 2.

and N-Back Difficulty was significant, but the expected three-way interaction between Subtraction Difficulty, Text-Entry Difficulty, and Interruption Duration did not reach significance. However, given that our model made specific *a priori* predictions, the overall ANOVA is overly conservative as it tests for any kind of three-way interaction, instead of the clear directional effects that the model predicted. We therefore subsequently performed simple effects analyses to investigate the model's prediction that we should only observe increasing costs with interruption duration in the *hard-hard* condition. We tested for the separate n-back and subtraction conditions whether there was an effect of Interruption Duration. These analyses confirmed that there was no effect of Interruption Duration as long as either the subtraction task or the n-back task was easy (all $Fs < 1$). In the *hard-hard* condition, on the other hand, we found a significant effect of Interruption Duration ($F(1,32) = 4.17$, $p = .0495$, $\eta_p^2 = .12$). Thus, resumption costs only increased with interruption duration in the *hard-hard* condition, which is in line with the predictions of the WMM theory.

Discussion. In the interruption experiments we tested the prediction of the WMM theory that interruptions should only lead to increasing resumption costs with interruption duration when both the primary and the secondary task require an intermediate representation. The two experiments confirm this prediction, and the model fits show that the WMM theory can account for the datasets.

In Interruption Experiment 2 we only observed increasing costs with interruption duration in the *hard-hard* condition, and no increasing costs in the *hard subtraction – easy n-back* condition. Thus, this experiment suggests that the increasing costs in Interruption Experiment 1 in the *hard text-entry – easy n-back* condition were caused by properties of the text-entry task, possibly by a dual-representation strategy. Therefore, the two experiments taken together suggest that interruptions only lead to increasing costs with interruption duration when both tasks require an intermediate representation. This contrasts with earlier findings (e.g., Hodgetts & Jones, 2006; Monk et al., 2008), and with the memory for goals theory that proposes that resumption costs always increase with the duration of an interruption (e.g., Altmann & Trafton, 2002).

Table 6.2 ANOVA results for resumption cost in Interruption Experiment 2.

Source	$F(1,32)$	p	η_p^2
Subtraction Difficulty	33.83	< .001	.51
N-Back Difficulty	25.38	< .001	.44
Interruption Duration	3.69	.064	.10
Subtraction × N-Back	5.44	.026	.15
Subtraction × Interruption Duration	1.77	.19	.05
N-Back × Interruption Duration	1.83	.19	.05
Subtraction × N-Back × Interruption	2.51	.12	.07

The main objective of conducting the interruption experiments was to investigate whether intermediate representations are processed via a declarative memory system with decay, as the WMM theory proposes. The experiments showed that when the primary task required a representation and was interrupted by a task that also needed an intermediate representation, the costs of resumption increased with interruption duration. This argues in favor of a declarative memory system with decay: the longer ago an intermediate representation was stored in declarative memory, the more its activation will have decayed, the longer it takes to retrieve and restore it and to resume the primary task. Furthermore, the WMM model matched the size of the effects (Figure 6.10C and 6.11B), while using the same parameters as in the model fits discussed above.

Summary

In the first major section of this paper, we reviewed experimental evidence that shows that the problem state resource can only maintain a single intermediate representation at a time. In this section, we investigated where intermediate representations are maintained when they are removed from the problem state resource. According to the WMM theory, they are processed via a separate declarative memory store in which items are subject to decay. In the first part of this section, we used a novel model-based fMRI analysis technique to show that activity related to the problem state resource on the one hand, and activity related to declarative memory processes on the other hand, are best represented in two different brain areas, implying two separate memory stores. In the second part of this section, we reported two interruption experiments that show that representations that are not maintained in the problem state resource are subject to decay. Thus, this combination of experiments implies that intermediate representations that cannot be maintained in the problem state resource are processed via a declarative memory store that is subject to decay, while representations in the problem state resource are not subject to decay.

In the next section we will look at the proposed strategic nature of using the problem state resource: according to the WMM theory the problem state resource, in combination with declarative memory, will only be used when it is faster than using the environment.

The Strategic Nature of using the Problem State Resource and the Environment

Using intermediate representations is only necessary when information is not available in the environment. For example, when solving a multi-column subtraction problem on paper, one would usually indicate on paper whether a carry is in progress, and thus use the environment as an external representation (e.g., Hollan et al., 2000; Kirsh, 1995; Wickens, 1992). Even when one is interrupted while solving a subtraction problem mentally, it is possible to reconstruct the representation using the previous columns after the interruption, instead of retrieving a representation from memory (as was required in the interruption experiments above). Based on the Soft Constraints Hypothesis (Gray & Fu, 2004; Gray et al., 2006), the WMM theory assumes that people always use the fastest strategy: whether that is using a representation in the problem state resource, retrieving a representation from declarative memory, or using the environment. In this section, we will present an experiment that supports this assumption.

For this experiment we again used the subtraction and text-entry dual-task described above, but added a condition in which on-screen support was provided for the subtraction task (see also Buwalda, Borst, Taatgen, & van Rijn, 2011). Figure 6.12 shows the interface of this condition. The '|' above the subtraction task indicates that currently no carry is in progress, it will turn into a '1' after a column that induced a carry. Thus, no intermediate representation is required in this Support condition for the subtraction task. The experiment had a factorial 2 × 2 × 2 within-subject design (Subtraction Difficulty (easy/hard) × Text-Entry Difficulty (easy/hard) × Support (yes/no)). Additional methods are reported in Appendix C.

Predictions

For this task, the WMM theory predicts the following. In the no-support condition of the subtraction task, intermediate representations cannot be reconstructed from the environment, because only a single column is shown at a time. In the support condition, on the other hand, the intermediate representation is shown on the screen. Thus, the costs of using the environment in the support condition consist of perceiving the support indicator. As long as the subtraction task is easy, no differences are predicted between the support and the no-support condition, as carries are never required. In the *hard subtraction – easy text-entry* condition, the model also expects the same behavior with and without support: Because a mental strategy is faster than perceiving the support indicator on the screen, the WMM theory predicts that subjects will not use the indicator in the support condition.

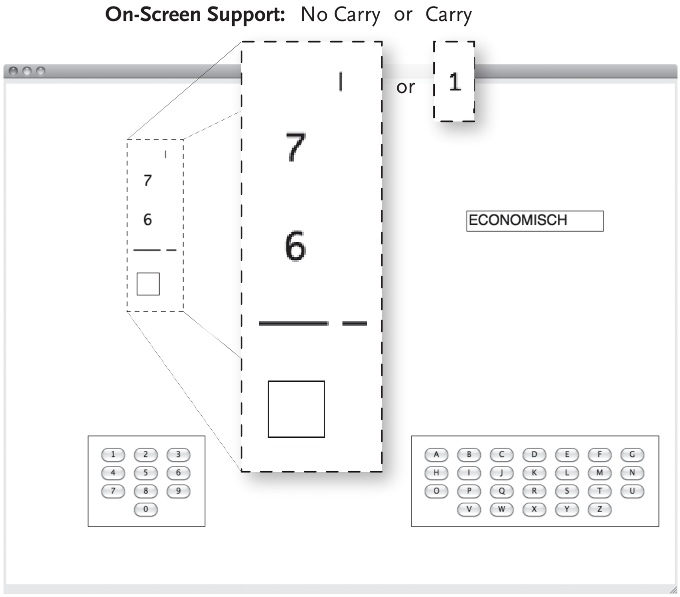


Figure 6.12 The interface of the experiment. The '|' in the subtraction task indicates that currently no carry is in progress, a '1' indicates a carry in progress.

Only in the *hard subtraction – hard text-entry* condition the WMM theory predicts a difference: in the support condition it is faster to use the environment than to use a declarative memory-based reconstruction (as used for all tasks above). However, using the environment in the *hard-hard* condition is still slower than using a mental representation in the *hard subtraction – easy text-entry* condition with support (because the support-indicator has to be perceived, which takes time). Therefore, in the support condition, the model expects slightly higher response times in the *hard-hard* condition than in the *hard subtraction – easy text-entry* condition. Thus, while an interaction effect between Subtraction Difficulty and Text-Entry Difficulty is expected in the no-support condition, as observed previously, a reduced interaction effect is predicted in the support condition.

While the WMM theory predicts interaction effects for response times both with and without support, for the accuracy data the model predicts that the interaction effect disappears with support. As the model uses the perfect information in the environment in the *hard-hard* condition with support, there is no reason for it to make more mistakes in this condition than in the *hard subtraction – easy text-entry* condition.

Results

Figure 6.13 shows the results of the experiment: the top panels show response times, the bottom panels accuracy. The complete ANOVA results are reported in Tables 6.3, 6.4, and 6.5. As the WMM theory predicted, subjects showed significant interaction

Table 6.3 Overall ANOVA results of the external support experiment, on the left for response times, on the right for accuracy.

	Response Times			Accuracy		
Source	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2
			Subtraction Task			
Support	8.02	.009	.26	65.7	< .001	.74
Subtraction	484.2	< .001	.95	103.8	< .001	.82
Text-Entry	26.3	< .001	.53	18.78	< .001	.45
Support × Subtraction	27.6	< .001	.55	66.8	< .001	.74
Support × Text-Entry	4.22	.05	.15	12.7	.002	.36
Subtraction × Text-Entry	29.35	< .001	.56	24.7	< .001	.52
Support × Sub. × Text-Entry	5.05	.03	.18	21.4	< .001	.48
			Text-Entry Task			
Support	2.85	.10	.11	3.20	.09	.12
Subtraction	105.5	< .001	.82	25.7	< .001	.53
Text-Entry	1.17	.29	.05	149.4	< .001	.87
Support × Subtraction	32.7	< .001	.59	< 1	–	–
Support × Text-Entry	3.96	.06	.15	3.87	.06	.14
Subtraction × Text-Entry	19.7	< .001	.46	27.1	< .001	.54
Support × Sub. × Text-Entry	9.29	.006	.29	< 1	–	–

Subtraction = Subtraction Difficulty, Text-Entry = Text-Entry Difficulty.

effects between Subtraction Difficulty and Text-Entry Difficulty both with and without support. Moreover, the interaction effect was smaller in the support condition, as shown by a significant three-way interaction effect between Subtraction Difficulty, Text-Entry Difficulty, and Support. This is in line with the predictions described above.

The accuracy data of the subtraction task show a similar pattern, except that the interaction between Subtraction Difficulty and Text-Entry Difficulty completely disappears in the Support condition. This is also in line with the model: while using the environment is slightly slower than using a mental strategy, it should not lead to more errors.

Model

As can be seen in Figure 6.13 (grey bars), the WMM model accounted well for the data. Both the effects on response times and accuracy of the subtraction task were captured. The size of the effects was also reflected by the model, which uses the same parameters as before.

Discussion

The experiment supports the assumption of the WMM theory that problem state resource usage is strategic: subjects always used the fastest option. As long as a mental

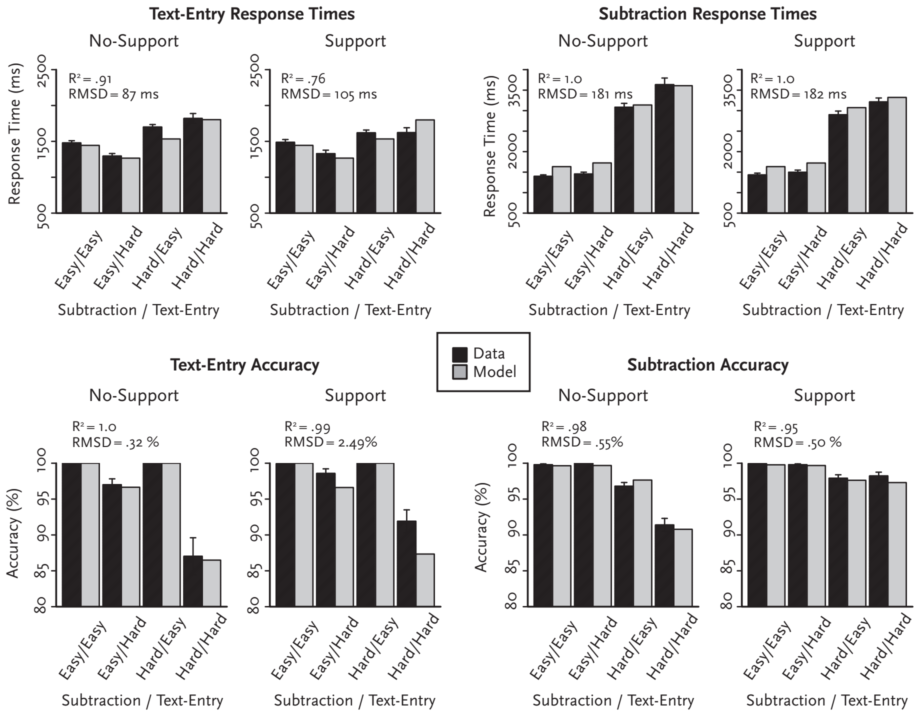


Figure 6.13 Results of the support experiment.

strategy was possible (up to the *hard subtraction – easy text-entry* condition with support), they seemed to prefer a mental strategy instead of using the environment, even when support was provided. In the *hard-hard* condition, when they had to use an intermediate representation for both tasks, they switched to environmental support when possible, resulting in faster response times with support than without support.

This is in accordance with the Soft Constraints Hypothesis (Gray & Fu, 2004; Gray et al., 2006), which states that humans adapt their behavior to minimize temporal costs, even if that leads to suboptimal behavior. That is exactly what we observed in the experiment, in which it might make more sense to always use the perfect knowledge in-the-world as opposed to imperfect knowledge in-the-head. While subjects were free to use the indicator in the Support condition – which seems to be the rational option, as it is always correct and requires less effort than remembering whether a carry was in progress – there was a clear difference between the *hard subtraction – easy text-entry* condition and the *hard-hard* condition, indicating that they used different strategies in these conditions. According to the WMM theory and the Soft Constraints Hypothesis this is because participants tried to minimize the temporal costs of the task, and therefore prefer the imperfect knowledge in-the-head over the perfect knowledge in-the-world.

Table 6.4 ANOVA results of the response time data; separate for Support and No-Support.

Source	RT No-Support			RT Support		
	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2
			Subtraction Task			
Subtraction Difficulty	357.9	< .001	.94	531.3	< .001	.96
Text-Entry Difficulty	22.0	< .001	.49	15.0	< .001	.40
Subtraction × Text-Entry	20.0	< .001	.47	14.5	< .001	.39
			Text-Entry Task			
Subtraction Difficulty	133.6	< .001	.85	46.4	< .001	.67
Text-Entry Difficulty	< 1	–	–	2.6	.12	.10
Subtraction × Text-Entry	26.5	< .001	.53	8.6	.007	.27

RT = response times.

Table 6.5 ANOVA results of the accuracy data; separate for Support and No-Support. Note that for the accuracy of the text-entry task we did not find any effects involving Support, which is why we collapsed over Support.

Source	Acc No-Support			Acc Support		
	$F(1,23)$	p	η_p^2	$F(1,23)$	p	η_p^2
			Subtraction Task			
Subtraction Difficulty	80.4	< .001	.77	36.8	< .001	.62
Text-Entry Difficulty	45.0	< .001	.66	< 1	-	-
Subtraction × Text-Entry	58.2	< .001	.72	1.1	.3	.05
			Text-Entry Task			
Subtraction Difficulty	25.9	< .001	.53			
Text-Entry Difficulty	173.0	< .001	.88		*Same as No-Support*	
Subtraction × Text-Entry	28.1	< .001	.55			

Acc =accuracy.

General Discussion

In this article we proposed the Working Memory in Multitasking theory, which describes how intermediate representations are processed in human multitasking. The WMM theory states that our cognitive system has a single-sized problem state resource for maintaining intermediate representations. When more than one representation is required for the tasks that make up the ‘multitask’, the representations that are currently not used are temporarily stored in a declarative memory store that is subject to decay. The use of multiple intermediate representations leads therefore to interference, because the representations have to be swapped in and out of the problem state resource via declarative memory. The WMM theory furthermore proposes that intermediate representations are processed mentally as long as that is faster than using the environment, but that otherwise the environment is used.

In the previous three sections we have discussed various data sets that support the WMM theory. First, we reviewed an experiment of Borst, Taatgen, and Van Rijn (2010) that showed that the problem state resource can maintain at most a single representation. We then turned to a novel fMRI analysis technique to support the assumption that discarded representations are stored in a separate memory store. In addition, with two new interruption experiments we showed that representations in this declarative memory store are subject to decay. Finally, we presented an experiment in which external support was given to the subjects, to show that intermediate representations are always processed in the fastest way possible, whether that is using the environment, the problem state resource, or declarative memory.

The main prediction of the WMM theory is that intermediate representations are an important factor in interference in human multitasking. This idea stems from the combination of the threaded cognition theory (Salvucci & Taatgen, 2008, 2011) and the ACT-R theory (e.g., Anderson, 2007). The data sets discussed in this article give therefore also indirect support for these theories. The WMM theory, being based on threaded cognition, is a multiple bottleneck theory (although we focused on the problem state bottleneck in this article). It therefore supports the idea of multiple separate bottlenecks, with interference effects depending on the requirements of the tasks at hand. According to the WMM theory one of those bottlenecks is the problem state resource. As it takes a relatively long time to restore intermediate representations to the problem state resource, this bottleneck causes substantial interference, both in response times and accuracy.

In the remainder of this article we will discuss the relation between the WMM theory and current working memory theories, its consequences for interruption studies, its relation to the Memory for Goals theory and the Soft Constraints Hypothesis, and finally real-world implications of the WMM theory.

Working Memory

The function of intermediate representations – temporarily maintaining information – is traditionally part of the concept of short-term or working memory (e.g., Baddeley, 1986; Baddeley & Hitch, 1974; Miller, 1956). However, where theories used to assume a limit of 7±2 items (Miller, 1956) or 4±1 items (Cowan, 2000), the WMM theory assumes a limit of only one directly accessible item. Other items that are traditionally part of working memory are in the WMM/ACT-R framework represented by highly activated items in declarative memory (Anderson, 2005; Daily et al., 2001; Lewis & Vasishth, 2005; Lovett et al., 2000). The combination of a single-sized problem state resource and highly active items in declarative memory gives a more traditional working memory capacity of around four to nine items (Anderson et al., 1998). In this framework, the representation in the problem state resource can be used immediately, while it takes time to retrieve and use the highly activated items in declarative memory.

Recent working memory theories have proposed similar ideas, in which the 'focus of attention' represents items in working memory that are directly available (e.g., Cowan, 1995; Garavan, 1998; Jonides et al., 2008; McElree, 2001; Oberauer, 2002, 2009;

Oberauer & Bialkova, 2009). For example, Oberauer (2009) proposed a model in which "declarative working memory" consists of activated items of long term memory, of which only a single item can be in the focus of attention. This idea of one directly available item and other highly active items in declarative memory constituting working memory is very similar to the concept of working memory in the WMM theory, as are other recent working memory theories (e.g., Jonides et al., 2008; McElree, 2001; Oberauer, 2002).

Interruptions

We illustrated the WMM theory with two interruption experiments. It is well known that interruptions of a task lead to a decrease in performance when recommencing this task (e.g., Gillie & Broadbent, 1989; McFarlane & Latorella, 2002). Two important factors that determine the severity of an interruption are the complexity and the duration of the interrupting task (e.g., Gillie & Broadbent, 1989; Hodgetts & Jones, 2006; Monk et al., 2008; Zijlstra et al., 1999). According to the WMM theory, these effects are partly caused by intermediate representations. As explained in detail above, whether the tasks need intermediate representations determines whether there is representation-related interference (complexity effect), while decay in declarative memory of intermediate representations causes the duration effect.

Interestingly, this leads to the prediction that the duration effect only plays a role when both tasks need an intermediate representation, as opposed to what is normally assumed (and in contrast to the Memory for Goals theory, e.g., Altmann & Trafton, 2002, see below). The presented experiments confirmed this prediction: as long as the primary task did not need an intermediate representation there was no effect of the duration of the interruption. Moreover, when the primary task required a representation, there was only an effect of interruption duration if the interrupting task also needed an intermediate representation.

While the WMM theory accounts for increasing costs with interruption duration, note that it does not (neither does it mean to) explain the basic interruption costs that we observed in all conditions (Figure 6.10 and 6.11). According to the WMM theory they cannot be ascribed to the processing of intermediate representations, but are caused by a different mechanism.

Memory for Goals

The Memory for Goals theory (Altmann & Trafton, 2002; Trafton et al., 2003) has been used to explain resumption costs in interruption tasks such as the ones presented in this article. It assumes that every task has an associated goal in declarative memory, and that when a task is interrupted, this goal starts to decay. When a task is recommenced after an interruption, the associated goal has to be retrieved from memory, which takes more time the longer the interruption was due to decay in memory (cf. Figure 6.4). Thus, Memory for Goals theory explains interruption costs by the time it takes to retrieve a goal from declarative memory.

Salvucci, Monk, et al. (2009) rephrased Memory for Goals in problem state terms: "which they [Altmann & Trafton, 2002] referred to as the *goal* [...] we call the *problem state*" (p. 799). Thus, they assume that every task has an associated intermediate representation in the problem state resource (instead of a goal), which has to be retrieved from declarative memory after an interruption. As support they presented a model fit of an interruption experiment by Monk et al. (2008). However, while Salvucci, Monk, et al. assumed that every task has an associated intermediate representation (which was true for the tasks they simulated), the WMM theory proposes that not all tasks need an intermediate representation. According to the WMM theory, only tasks with associated intermediate representations will lead to increasing resumption costs with interruption duration, thereby explaining the flat lines in the easy conditions in the interruption experiments described above.

Strategic Behavior and the Soft Constraints Hypothesis

The WMM theory assumes that humans always use the fastest strategy, whether that is using information in-the-head or using information in-the-world. This assumption is based on the Soft Constraints Hypothesis, which proposes that behavior is adapted to a task by minimizing temporal costs (Gray & Fu, 2004; Gray et al., 2006). The support experiment presented in this article yields additional behavioral evidence for this hypothesis: while it might have been expected that subjects always use external support when it is available, this did not seem to be the case. An open question is how subjects decided which strategy to use: how did they learn that it was faster to use information in-the-head than to use the information presented in the environment?

To discover the best strategy, Gray et al.'s (2006) Ideal Performer Model used reinforcement learning (e.g., Sutton & Barto, 1998). However, they "make no claim that the process followed by the [reinforcement learning] algorithm mimics any process followed by human cognition" (Gray et al., 2006, p. 465). Recently, Janssen and Gray (in press) investigated whether reinforcement learning could also provide a cognitively plausible explanation of the data. They concluded that reinforcement learning could be used to simulate the human data, especially if it used 'time' as the reward parameter. This is in line with other efforts to use reinforcement learning to explain human behavior (see, e.g., Daw & Frank, 2009, for a special issue on reinforcement learning and higher-level cognition). As the subjects in our experiment received sufficient practice, it is possible that reinforcement learning can also be used to explain the strategic choices in our dataset.

Real World Implications

In this article we have presented several behavioral experiments that highlight the interference effects of the WMM theory. While these interference effects are relatively large, one could wonder if the processing of intermediate representations also plays a role in everyday life. First evidence that this is the case comes from an experiment in which subjects had to keep a (very basic) simulated car in the middle of the road

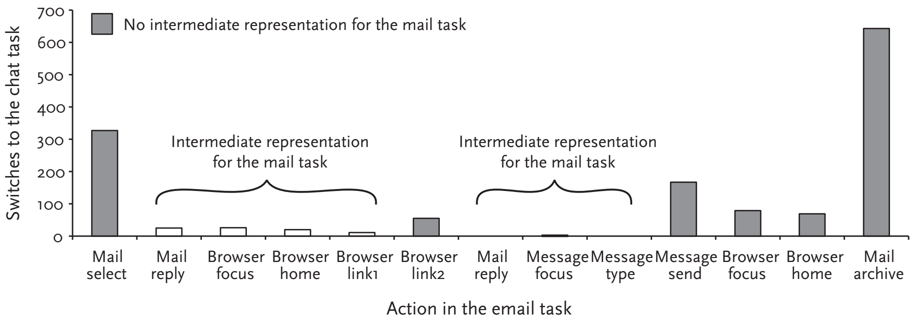


Figure 6.14 Results of the Salvucci and Bogunovich (2010) experiment. Adapted from Figure 3 (Salvucci & Bogunovich, 2010).

while entering addresses into a navigation device (Borst & Taatgen, 2007). Both tasks had two conditions: either they needed an intermediate representation or not. When both tasks required an intermediate representation, performance was slower than in all other conditions.

More recently, Salvucci and Bogunovich (2010) let subjects perform a customer-support task in which the use of intermediate representations was manipulated. Subjects in their experiment had to reply to emails inquiring about the price of a certain product. These prices had to be looked up on a simulated internet. During certain parts of this task, subjects had to maintain an intermediate representation (white bars in Figure 6.14). While replying to the emails and looking up the information on the internet, subjects additionally had to perform a chat task: Sometimes a chat window was highlighted, indicating that a new message had arrived to which the subjects had to respond. Subjects were free to choose when they switched to the chat task. Figure 6.14 shows the results: they switched almost exclusively to the chat task when no intermediate representation had to be maintained (the grey bars). This indicates that the subjects were aware that there is a cost associated with using intermediate representations, and thus that multitasking interference due to the use of intermediate representations has an impact on real-world tasks.

Conclusion

In this article we proposed the Working Memory in Multitasking theory. This theory states that intermediate representations are an important factor in multitasking interference. As support for the WMM theory we have reviewed experiments from the literature, and in addition tested predictions of the theory in three new experiments, showing that it accounts for data ranging from concurrent tasks to interruption studies, and from laboratory experiments to more real-world tasks.

Appendix A: Additional Methods of Interruption Experiment 1 and 2

In this appendix we describe the subjects, stimuli, procedure, and statistical procedure of Interruption Experiments 1 and 2.

Interruption Experiment 1: Text-Entry and N-Back

In this experiment subjects had to perform a text-entry task, which was twice every trial interrupted by an n-back task. The design of the task is discussed in the main text.

Subjects

16 students of the University of Groningen participated in the experiment for course credit (12 female, age range 18–22, mean age 19.2). All subjects had normal or corrected-to-normal visual acuity. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

Stimuli and Apparatus

The 10 letter words for the hard version of the text-entry task were handpicked from a list of high-frequency Dutch words (CELEX database; Baayen et al., 1993) to ensure that similarities between words were kept at a minimum. These stimuli were also used in the easy text-entry task, except that the letters within the words were scrambled (under the constraint that a letter never appeared twice in a row). Thus, subjects were presented pseudo-random sequences of letters that they had to enter one-by-one in the easy condition. By scrambling the words, we controlled for letter-based effects, while preventing the use of strategies to predict the next letter.

The n-back stimuli were generated during the experiment. Subjects had to respond 'same' in 50% of the cases (thus, in the easy, 1-back version, the digit was the same as the previous digit, in the 2-back version the same as the digit 2 back), and 'different' in the other 50% of the cases. Additionally, no more than three digits in a row could be the same. Subjects had to press 'x' or 'z' on the keyboard to indicate 'same' or 'different'. Whether 'x' or 'z' indicated 'same' or 'different' was counter-balanced over subjects.

Procedure

Each trial in the experiment started with presenting the conditions of the text-entry task and the n-back task on the screen, for example: N-Back: easy, Text-Entry: hard. This was presented for three seconds, after which the text-entry task was shown on the screen. After entering between 2 and 6 letters the first n-back interruption was started. No warning was given, but instead of the next text-entry letter the n-back task was shown (Figure 6.9). The subject had to do 3, 6, or 9 n-back steps (balanced over the text-entry and n-back conditions). Then the text-entry interface was shown again. After another 2 to 6 letters the second n-back interruption started, after which the subject could finish the text-entry task. When the subject had entered the complete 10-letter word, feedback was presented, indicating the number of correct letters in the text-entry task. The text-entry feedback was presented for 3 seconds, followed by a fixation screen for 4 seconds. Afterwards the next trial started. Feedback was continuously presented for the n-back task: every time an incorrect or no response was given, a buzzer was sounded.

The experiment consisted of a practice block and two experimental blocks. The practice block started with three trials of only the easy text-entry task, followed by three trials of the hard text-entry task, seven 6-step trials of the easy n-back task, and seven 7-step trials of the hard n-back task. Then the real task was practiced in four trials. The seven n-back trials (per condition) increased in speed in four steps: in the first trial the digit was presented for 1700 ms followed by a 733 ms mask, in the second and third trial they were presented for 1500/533 ms, in the fourth and fifth trials for 1300/383 ms, and in the sixth and seventh trials for 1100 ms with a 233 ms mask (as in the real experiment).

The two experimental blocks were the same, each consisted of 36 trials: 2 (easy/hard text-entry) × 2 (easy/hard n-back) × 3 (3/6/9 steps first interruption) × 3 (3/6/9 steps second interruption). The order of these conditions was randomized within a block. The interruption points in each trial were also randomized: the first interruption came after the 2nd up to the 6th letter, the second interruption between the 4th and the 8th letter. The distance between the interruptions was a minimum of 2 letters, resulting in 15 different combinations. The complete experiment consisted of 72 experimental trials, and lasted for about 90 minutes. Between the blocks subjects could take a short break.

Statistical Procedure

We only analyzed the data from the experimental phase of the experiment. A response time in the text-entry task was defined as the time between two responses, or, directly after an interruption, as the time between the reappearance of the text-entry task and the response. First responses of each trial were removed. Outliers were removed from the data (RTs < 250 ms or > 10,000 ms), after which we removed data exceeding three standard deviations from the mean per condition per subject (in total 0.92% of the data was removed). All F- and p-values are obtained from repeated measure ANOVAs; all error bars depict standard errors. Analyses on response times are only for correct responses.

Interruption Experiment 2: Subtraction and N-Back

In this experiment subjects had to perform a subtraction task, which was twice every trial interrupted by an n-back task. The design is similar to the design of Interruption Experiment 1, except when noted otherwise.

Subjects

39 students of the University of Groningen participated in the experiment for course credit. Three subjects had to be excluded because of not being able to do the hard subtraction task (< 65% columns correct or slower than 3,500 ms per response), two subjects for not doing the hard n-back task (< 80% correct), and one subject had to be excluded because of health problems, leaving 33 complete data sets (23 female, age range 18-27, mean age 20.7). All subjects had normal or corrected-to-normal visual acuity. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

Design

Instead of text-entry, solving 10-column subtraction problems was the primary task in the second interruption experiment. Responses had to be made using the numeric keypad of the keyboard. The n-back task was again used as the interrupting task, but now with letters instead of digits.

Stimuli and Apparatus

The stimuli for the subtraction task were generated anew for each subject. The subtraction problems in the hard version always featured five carries, and resulted in 10-digit answers. The n-back stimuli were again generated on the fly, using the same constraints as in Interruption Experiment 1.

Statistical Procedure

We only analyzed the data from the experimental phase of the experiment. A response time in the subtraction task was defined as the time between two responses, or, directly after an interruption, as the time between the reappearance of the subtraction task and the response. First responses of each trial were removed. Outliers were removed from the data (RTs < 250 ms or > 15,000 ms), after which we removed data exceeding three standard deviations from the mean per condition per subject (in total 2.74% of the data was removed). All *F*- and *p*-values are obtained from repeated measure ANOVAs; all error bars depict standard errors. Analyses on response times are only for correct responses.

Appendix B: Effects of a Dual-Representation Strategy on Previous Results

In this appendix we report the effects of a dual-representation strategy for the text-entry task on previously published results. According to the results of Interruption Experiment 1 (see the main text), it is likely that two intermediate representations are used for the text-entry task, instead of one. Thus, when a word has to be entered in the hard text-entry task, one representation is used for storing the word ('information') and one for the position in the word ('6th letter'). Such a model implementation resulted in a good fit for Interruption Experiment 1. However, if a dual-representation strategy is indeed the correct way of explaining the results of the text-entry task, the question is whether previously published model fits still hold. To investigate this we implemented the dual-representation strategy for Experiment 1 of Borst, Taatgen, and Van Rijn (2010) and for the model-based analysis reported in Borst et al. (2011).

Borst, Taatgen, and Van Rijn (2010)

The design of Experiment 1 (Borst, Taatgen, & Van Rijn, 2010) is reported in "Multiple Intermediate Representations cause Multitasking Interference". Figure 6.6 in the main text shows the original model fit, Figure 6.15 the model fit with a dual-representation strategy. The effect of a dual-representation strategy for the text-entry task is small: the costs in response times and accuracy for the text-entry task are slightly over-estimated. This is caused by the model now having to swap out two intermediate representations for the text-entry task throughout the experiment. In the *easy subtraction – hard text-entry* condition this leads to slightly higher response times, because the new model has to retrieve an intermediate representation from memory on each step of a trial, even in this condition. However, most of these costs are absorbed in the normal costs of doing the task (i.e., perceptual and motor costs). In the *hard-hard* condition the model now has to retrieve two intermediate representations from memory at each step of a trial, leading to similar interaction effects as observed before. Taken into account that we did not refit the model parameters, this shows that a dual-representation strategy can account for the data of Experiment 1 of Borst, Taatgen, & Van Rijn.

Model-Based fMRI: Borst, et al. (2011)

In "Neuroimaging Evidence for Two Separate Memory Stores" we discussed a model-based fMRI analysis that showed that the problem state resource and declarative memory are best represented by two different brain areas (Figure 6.8, main text). We now reanalyzed the data with a dual-representation model. Figure 6.16 shows the results. The analysis still found two separate regions for the problem state resource and declarative memory. This indicates that a dual-representation model is also compatible with the brain results we found before, which, in turn, correspond to the standard mapping between ACT-R modules and brain regions (e.g., Anderson, 2007).

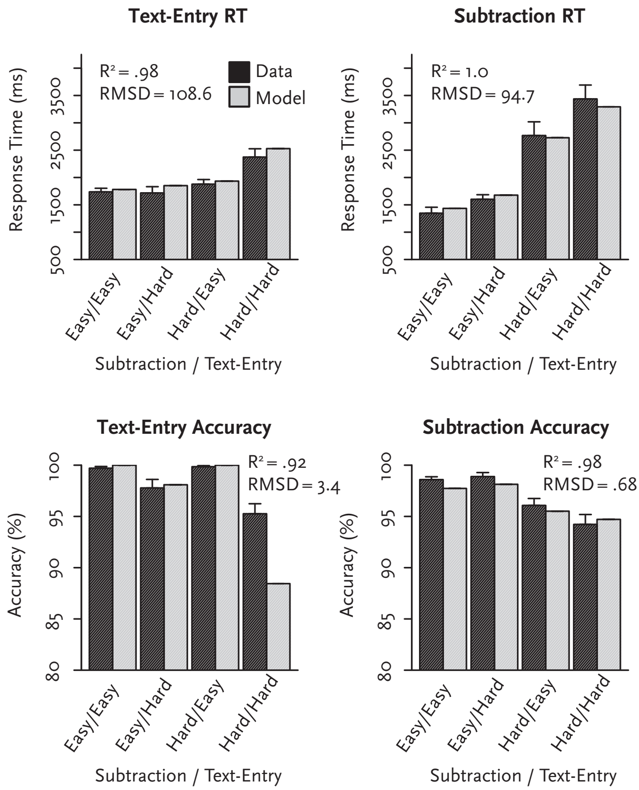


Figure 6.15 Results of Experiment 1 with a dual-representation model. Error bars represent standard errors. RMSD = root-mean-square deviation; RT = response time.

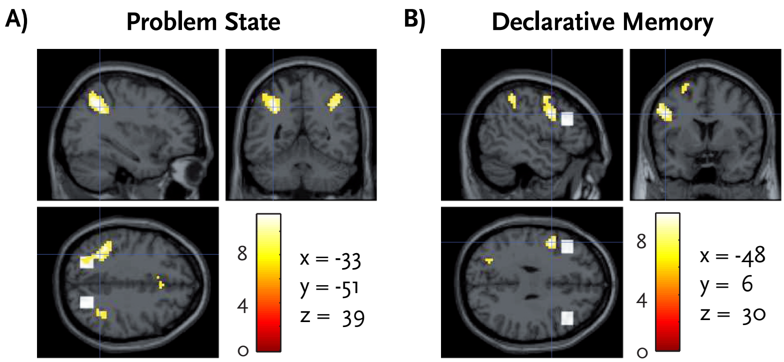


Figure 6.16 Results of the model-based fMRI analysis with a dual-representation model. The best fitting areas for a) the problem state resource and b) declarative memory. The white squares indicate ACT-R's predefined regions (e.g., Anderson et al., 2008), xyz-coordinates are of the most significant voxel.

Appendix C:

Additional Methods of the Support Experiment

In this appendix we describe the subjects, design, and statistical procedure of the experiment described in the section 'The Strategic Nature of using the Problem State Resource and the Environment.' In this experiment subjects had to alternate between a subtraction task and a text-entry task (as in Borst, Taatgen, & Van Rijn, 2010). Both tasks were presented in two versions: an easy version in which there was no need to maintain a intermediate representation, and a hard version in which participants had to maintain an intermediate representation from one response to the next. In addition, in one condition external support was displayed on the screen for the subtraction task. The general design of the task is discussed in the main text.

Subjects

33 students of the University of Groningen participated in the experiment for course credit or monetary compensation of €10. Four participants were rejected because they scored less than 75% correct where the other participants scored >95% correct. Two subjects were rejected because they did not adhere to task instructions, and three because of recording problems of the eye tracker. This leaves 24 complete datasets (17 female, age range 18-43, mean age 20.5). All subjects had normal or corrected-to-normal visual acuity. Informed consent as approved by the Ethical Committee Psychology of the University of Groningen was obtained before testing.

Stimuli and Apparatus

The stimuli for the subtraction task were generated anew for each subject. The subtraction problems in the hard version always featured six carries, and resulted in 10-digit answers. The 10 letter words for the hard version of the text-entry task were handpicked from a list of high-frequency Dutch words (CELEX database; Baayen et al., 1993) to ensure that similarities between words were kept at a minimum. These stimuli were also used in the easy text-entry task, except that the letters within the words were scrambled (under the constraint that a letter never appeared twice in a row). Thus, participants were presented pseudo-random sequences of letters that they had to enter one-by-one in the easy condition. By scrambling the words, we controlled for letter-based effects, while preventing the use of strategies to predict the next letter.

The experiment was presented full screen on a 20.1" monitor. Participants were sitting at a normal viewing distance, about 70 cm from the screen.

Procedure

Each trial started with the presentation of a fixation cross for 6 seconds. The fixation cross was followed by two horizontally aligned colored circles representing the tasks.

The color of the circles indicated the difficulty levels of the tasks (on the left for the subtraction task, on the right for the text-entry task; green for easy, red for hard). The circles stayed on the screen for 1 second, followed by a fixation cross for 600 ms, after which the subtraction and text-entry tasks appeared. Subjects had to begin with the subtraction task, and then alternate between the two tasks. After completing both tasks, a feedback screen was shown for 2 seconds, indicating how many letters/digits were entered correctly. Before the next trial started, a fixation screen was shown for 2 seconds.

The experiment consisted of a practice block and two experimental blocks. One of the experimental blocks contained the support condition; the order was counter-balanced over participants. The practice block consisted of 12 single task trials (4 subtraction trials with 10 columns visible, 4 subtraction trials with one column visible, and 4 text-entry trials), followed by a block of 4 multitasking trials: all combinations of subtraction and text-entry (*easy-easy*, *hard-easy*, *easy-hard*, and *hard-hard*). Both experimental blocks consisted of 28 multitasking trials. Before the second block the subtraction task was practiced again, to familiarize the participants with using the carry indicator if they did not use this in the first block, or with performing the task without the indicator in the other case. Subtraction and text-entry conditions were randomized within a block. The complete experiment consisted of 56 experimental trials, and lasted for about 90 minutes. In between blocks participants could take a short break.

Statistical Procedure

We only analyzed the data from the experimental phase of the experiment. A response time in the subtraction task is defined as the time between a response in the text-entry task and a response in the subtraction task; a response time in the text-entry task as the time between a response in the subtraction task and a response in the text-entry task. First responses of each trial were removed. Outliers were removed from the data (RTs < 250 ms or > 10,000 ms), after which we removed data exceeding three standard deviations from the mean per condition per participant (in total 2.2% of the data was removed). All *F*- and *p*-values are obtained from repeated measure ANOVAs; all error bars depict standard errors. Accuracy data were transformed using an arcsine transformation before being submitted to the ANOVA.

Summary & Concluding Remarks

In which we briefly summarize the findings presented in this dissertation and finish with some concluding remarks.

7

Chapter

Summary & Concluding Remarks

7

Chapter

Summary & Concluding Remarks

I started this dissertation discussing an application that makes it harder to multitask: Concentrate[1]. According to the application's website, by enforcing monotasking it "helps you work and study more productively". A possible reason why Concentrate might make you more productive is the topic of this dissertation: the problem state bottleneck. In the preceding chapters we have shown that, due to this bottleneck, having to maintain multiple intermediate representations at the same time leads to a decrease in performance, both in time and accuracy. By focusing on a single task – for example with the help of Concentrate – it is more likely that you use at most one intermediate representation, resulting in better overall performance. In this last chapter, I will briefly summarize our findings and the resulting theory, and along the way discuss some high-level implications (more detailed discussions of the results can be found in the previous chapters). This chapter is organized around the different methodologies that we applied: behavioral experiments, cognitive modeling, pupil dilation, and neuroimaging.

Behavioral Results

First evidence for a problem state bottleneck came from the three behavioral experiments that we discussed in Chapter 2. In the first experiment, by varying the use of intermediate representations and thereby the use of the problem state resource, we showed that performance decreased considerably when more than one representation was required at a time. The second and third experiment were carried out to show that this decrease in performance was not caused by an effect of memory load or by a phonological loop bottleneck, respectively. In Chapter 3, we extended the support for a problem state bottleneck by showing that the interference was indeed due to the use of intermediate representations: when such a representation was presented in the environment, performance suffered less than when a representation had to be maintained mentally. These results run counter to classical working memory theories that assume that people can maintain up to 7±2 items in short-term memory (Miller, 1956). However, they match more recent theories, which propose a very limited focus of attention in working memory of only one or two items that can be used without a time cost (e.g., Cowan, 1995; Garavan, 1998; Jonides et al., 2008; McElree, 2001; Oberauer, 2002, 2009).

Following these results, in Chapter 6 we took a closer look at what happens to intermediate representations that cannot be maintained in the problem state resource. According to the underlying theories, ACT-R and threaded cognition, these representations are stored in a declarative memory store in which their memory strength decays over time. To test this assumption, we conducted two so-called interruption experiments. In these experiments, a primary task was interrupted for varying durations by a secondary task. The data showed that (1) when both tasks needed

[1] *http://getconcentrating.com/*

an intermediate representation the time to resume the primary task was higher than in the other conditions, and (2) that in that condition the time to resume the primary task increased with interruption duration. This matched the predictions of our theory, and indicated that intermediate representations are indeed stored in a declarative memory store when they cannot be maintained in the problem state resource.

Cognitive Models

To account for the results of the experiments we developed computational cognitive models. These models were instantiated in the cognitive architecture ACT-R (e.g., Anderson, 2007). To simulate the multitasking aspects of the experiments we used threaded cognition theory (Salvucci & Taatgen, 2008, 2011). On the one hand, the models show that the observed results can indeed be explained by a bottleneck in the problem state resource. On the other hand, the modeling results also cross-validate the ACT-R theory and threaded cognition. First, our models add new tasks to the expanding set of data ACT-R can account for (see also http://act-r.psy.cmu.edu/), and show thereby that ACT-R is a plausible psychological theory and at the same time a useful modeling framework. With respect to threaded cognition, the modeling accounts show that a multiple-bottleneck theory can explain our datasets, while this would have been difficult with a single central bottleneck account (e.g., Pashler, 1994). For example, while we mostly focused on the problem state bottleneck, to explain the effects of the listening task (Experiment 3, Chapter 2) the bottleneck in declarative memory was crucial. This also shows the added value of using a cognitive architecture: the different effects of, and the interactions between, the problem state resource (Experiment 1 & 2, Chapter 2), the visual resource (Chapter 3), and declarative memory (Experiment 3, Chapter 2; the interruption experiments, Chapter 6) were all necessary to explain the data. The effects of the different resources automatically follow from using a cognitive architecture, because the resources were implemented and supported previously (Anderson, 2007; Cooper, 2007; Kieras & Meyer, 1997; Newell, 1990; Van Maanen et al., 2009). If one would want to model our datasets with 'single-issue models', many more ad-hoc assumptions would have to be made, resulting in weaker modeling accounts.

Pupil Dilation Results

To investigate whether the decrease in performance due to the problem state bottleneck was accompanied by an increase in mental workload, we measured people's pupil dilation in Chapter 3. In addition to the interaction effects in the response time and accuracy data, we indeed also observed an over-additive interaction effect in pupil size, with most dilated pupils in the conditions where more than one intermediate representation had to be used. As it is well known that pupil dilation reflects mental workload (for reviews, see e.g., Beatty, 1982; Steinhauer & Hakerem, 1992), we interpreted this as evidence for increased mental effort when the limits of the problem state resource were reached. While it is still unclear what exactly is reflected by the pupillary response, these results could indicate that one factor is the use or maintenance of intermediate representations.

Neuroimaging Results

As the behavioral, modeling, and pupil dilation results all indicated a problem state bottleneck, the next step was to locate the neural correlates of this bottleneck. To this end we conducted two fMRI experiments (Borst et al., 2009, and Chapter 4 & 5). To analyze the fMRI data we first applied a Regions-of-Interest technique that is commonly used in combination with ACT-R (e.g., Anderson, 2007; Anderson et al., 2008). We used this method to see whether our model was able of making *a priori* predictions of the fMRI data in various regions in the brain, to thereby validate the model (Borst et al., 2009 and Chapter 4). In general the model's predictions were accurate, indicating that the model captures important aspects of the data. However, a discrepancy between the model's predictions and the data was found in the posterior parietal cortex, a region that is associated with the problem state resource. There are at least two possible reasons for this discrepancy: the model's assumptions could be incorrect or the ACT-R–brain link could be incorrect or incomplete.

Because the behavioral data in combination with the model delivered strong support for the problem state bottleneck, we focused on how we could improve the ACT-R–brain link. First, based on the broader fMRI literature and an unpublished dataset (Kao & Anderson, personal communication), we hypothesized that the area in the parietal cortex not only reflects problem state actions, but also visual-spatial activity. When we added this hypothesis to the model, the model-data match improved considerably (Borst et al., 2009). As this is based on a single experimental paradigm it cannot yet be used to validate other models. However, it does warrant deeper investigation, to further improve the mapping between ACT-R and the brain.

The results above are all based on the existing mapping between ACT-R's resources and brain areas. However, this naturally means that the results depend on the correctness of this predefined mapping. To investigate where the resources of our model are best represented in the brain – and to see whether that corresponds to the predefined mapping – we applied a novel model-based fMRI analysis technique in Chapter 5. This analysis technique, in which we regressed the model's predictions directly against the fMRI data, showed that the predictions of the problem state resource mapped best onto a region in the posterior parietal cortex that is slightly anterior to ACT-R's predefined region (in general, the results were surprisingly consistent with the ACT-R mapping). The measured BOLD response in this region was much closer to the model predictions than the response in the predefined region. However, the fit was still not perfect, meaning that either the model has to be improved, or the connection between the model and the brain (for example by adding visual-spatial activity, as implied above). One way to investigate this would be by entering all model resources in a single linear model instead of in several separate linear models (as we did in Chapter 5). By using one linear model, it is possible to have a combination of model resources predict

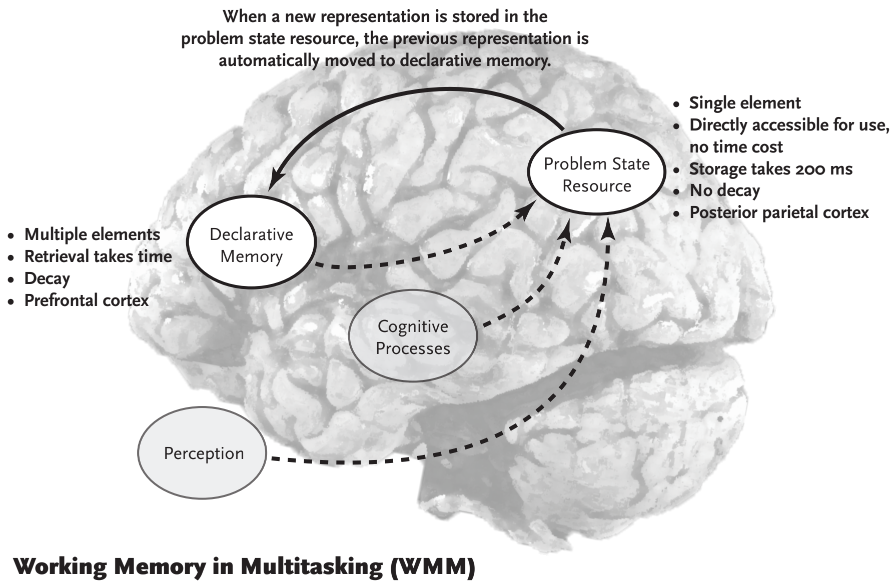


Figure 7.1 Overview of the Working Memory in Multitasking theory.

activity in one brain area. However, for that analysis to work we first need a better experimental paradigm, which dissociates all different resources of the model.

Besides yielding a better mapping between ACT-R and the brain, the model-based fMRI analysis method also leads to more precise function-brain mappings than traditional fMRI analysis methods. Traditionally, fMRI data is regressed against the different conditions of an experiment. The results of those regressions are then subtracted from each other, in principle isolating a certain cognitive function. However, by using a cognitive model – especially one developed in a cognitive architecture – this link between function and data is much more direct, as now a very well described (computationally implemented and previously validated) function is directly regressed against the brain data. While this does not completely solve the problem of fMRI data only showing that "the mind happens [...] north of the neck" (Fodor, 1999), the combination of computational models and fMRI data at least brings us one step closer to understanding the mind.

Conclusion: Working Memory in Multitasking

In Chapter 6 we introduced our over-arching theory: Working Memory in Multitasking (WMM). As this theory is based on all results presented in this dissertation, you could say that it is *the* conclusion of this dissertation. In that sense, Figure 7.1 is *the* summary of this dissertation. As shown in Figure 7.1, the core of the WMM theory is a single-sized problem state resource, which leads to interference when it has to be used by multiple tasks at the same time. When an intermediate representation is removed from the problem state resource it is automatically stored in declarative memory, where it starts to decay. This means that the longer a representation cannot be used in a multitasking situation, the harder it will be to retrieve it and resume an interrupted task. To support the WMM theory we have presented behavioral (Chapter 2 & 3), pupil dilation (Chapter 3), fMRI (Chapter 4 and 5), and computational cognitive modeling support (all chapters). To me, this is the essence of cognitive science: using formal methods in an interdisciplinary manner to investigate and understand the human mind.

Samenvatting

Deze samenvatting, en ook grote delen van dit proefschrift, zijn geschreven met behulp van het programma 'Concentrate'[1]. Concentrate is geen normaal programma, het kan niet gebruikt worden om mee te chatten, te schrijven, of te internetten. Nee, Concentrate zorgt er juist voor dat er niet teveel dingen tegelijkertijd gedaan kunnen worden, het dwingt *monotasking* af. Want terwijl in onze samenleving *multitasken* langzamerhand tot de standaard wordt verheven – hoe vaak zie je wel geen bellende fietsers? – hebben wetenschappers juist laten zien dat mensen vaak minder goed presteren als ze meerdere dingen tegelijkertijd doen.

Dit proefschrift gaat over één van de oorzaken van multitaskingproblemen: een beperking in het verwerken van tussenresultaten (bijvoorbeeld '$3x = 12$' wanneer '$3x - 7 = 5$' opgelost moet worden) . We laten zien dat als mensen een tussenresultaat voor meerdere taken tegelijkertijd moeten onthouden, hun prestatie op deze taken sterk vermindert. Volgens onze theorie worden de tussenresultaten namelijk opgeslagen in de zogeheten werkgeheugen-module in ons brein. Deze module kan echter maar één tussenresultaat tegelijkertijd opslaan, en fungeert dus als een beperkende factor in het multitasken wanneer mensen meerder tussenresultaten nodig hebben. Als dat het geval is moeten de tussenresultaten namelijk constant uitgewisseld worden tussen de werkgeheugen-module en het gewone geheugen. Deze uitwisseling kost tijd en kan soms misgaan, en leidt daarom tot een verminderde multitaskingprestatie.

In deze samenvatting zal ik proberen duidelijk te maken hoe we deze beperking in het verwerken van tussenresultaten (door ons de *problem state bottleneck* genoemd) onderzocht hebben, en wat de specifieke resultaten waren. Ik zal eerst de onderzoeksmethoden toelichten die we gebruikt hebben, gevolgd door de resultaten en onze uiteindelijke theorie.

Onderzoeksmethoden

Gedragsexperimenten

We hebben de *problem state bottleneck* met een aantal verschillende methoden onderzocht. Ten eerste hebben we gebruik gemaakt van gedragsexperimenten. In deze experimenten hebben we proefpersonen verschillende taken laten uitvoeren achter de computer, en gekeken hoe de eigenschappen van de taken hun prestaties beïnvloedden. Hierbij hebben we gekeken naar reactietijden en naar fouten die proefpersonen maken (bv. in Hoofdstuk 2), maar ook naar wat het effect van de taken was op de pupilgrootte van de proefpersonen (Hoofdstuk 3). Sinds de jaren zestig is het namelijk bekend dat de grootte van onze pupillen beïnvloed wordt door hoe moeilijk een taak voor ons is: hoe moeilijker de taak, hoe groter onze pupillen. Door het meten van pupilgrootte is het dus mogelijk om uit te vinden hoe moeilijk een taak is, zonder dat het de proefpersonen expliciet gevraagd hoeft te worden en zonder dat de taak onderbroken

[1] *http://getconcentrating.com/*

wordt. Naast deze standaard psychologische onderzoeksmethoden hebben we echter ook gebruik gemaakt van twee methoden die waarschijnlijk iets meer uitleg behoeven: cognitief modelleren en modelgebaseerde neurowetenschap.

Cognitieve Modellen

Na het doen van een aantal gedragsexperimenten, is het relatief gemakkelijk om een theorie te bedenken die de resultaten verklaart. Stel dat proefpersonen altijd meer fouten maken als ze twee tussenresultaten moeten onthouden in plaats van één. Dat zou uitgelegd kunnen worden met een theorie die stelt dat we een werkgeheugenmodule hebben die maar één enkel tussenresultaat tegelijkertijd kan opslaan. Maar wat betekent dat? Vergeten we het andere tussenresultaat dan? Als blijkt dat proefpersonen de taak in de helft van de gevallen nog wel goed doen, maar veel trager zijn dan in een situatie met maar één tussenresultaat, dan kan de theorie uitgebreid worden met het idee dat het tweede tussenresultaat opgeslagen wordt in ons normale geheugen. Omdat het ophalen uit ons normale geheugen meer tijd kost leidt dat tot tragere reacties, en soms zal een tussenresultaat vergeten worden. Maar... hoeveel tijd kost het dan om een tussenresultaat op te halen, en hoe vaak wordt het vergeten?

Zoals hopelijk duidelijk is geworden uit dit voorbeeld zijn verbale theorieën vaak niet precies genoeg beschreven om voor een goede uitleg te zorgen. Aan de ene kant is het vaak onduidelijk of een verbale theorie eigenlijk wel leidt tot de gevonden resultaten (wat betekent 'langzamer'?), en aan de andere kant kan een verbale theorie vaak op meerdere manieren geïnterpreteerd worden, en kan de theorie daarom niet goed getest worden in een experiment. Vanuit een behoefte aan meer precieze theorieën zijn psychologen daarom computationele cognitieve modellen gaan ontwikkelen. Een computationeel cognitief model is niets anders dan de implementatie van een psychologische theorie als een computerprogramma. Dit programma kan dan gebruikt worden om gedrag te simuleren. Aan de ene kant dwingt dit af dat alle details van een theorie expliciet gemaakt worden – anders werkt het programma simpelweg niet – en aan de andere kant maakt het precies duidelijk welke voorspellingen de theorie doet.

In dit proefschrift heb ik gebruik gemaakt van modellen die de complete taak kunnen uitvoeren. Dat wil zeggen dat het model dezelfde interface 'ziet' als de proefpersonen, en ook antwoord moet geven met behulp van een virtueel toetsenbord en een virtuele muis. Als het ware zijn deze modellen dus virtuele proefpersonen die één of meerdere taken kunnen uitvoeren. Dit betekent dat we direct de prestaties (bijvoorbeeld reactietijden en fouten) van onze modellen konden vergelijken met die van onze echte proefpersonen, en zo konden kijken waar onze theorie klopte en vooral waar de theorie nog verbeterd moest worden.

Hoewel een cognitief model al een stuk beter is dan een verbale theorie, bestaat er nog steeds het gevaar dat een model alleen maar werkt voor de taak waar het voor ontwikkeld is. Het kan bijvoorbeeld zijn dat het geheugen van een model allerlei eigenschappen wordt toegedicht die wel werken voor de huidige taak, maar niet voor andere taken. Om dit probleem te voorkomen zijn alle modellen in dit proefschrift geïmplementeerd in een cognitieve architectuur. Kortgezegd is een cognitieve architectuur een verzameling cognitieve modellen die sterk op elkaar lijken, en gebruikt worden om data van meerdere verschillende taken te verklaren. Voor de specifieke taak waarin een onderzoeker geïnteresseerd is moeten dan nog wel de details van het model geïmplementeerd worden, maar daarbij kan gebruik gemaakt worden van (bijvoorbeeld) een geheugensysteem dat al gebruikt is om andere taken te verklaren. Hierdoor wordt voorkomen dat modellen alleen maar de resultaten van één specifieke experiment verklaren.

In dit proefschrift heb ik gebruik gemaakt van de cognitieve architectuur ACT-R, ontwikkeld door John Anderson. In het verleden is ACT-R gebruikt om taken te modelleren die variëren van simpele reactietijd-taken tot autorijden en luchtverkeersleiding. Voor het modelleren van onze experimenten hebben we bijvoorbeeld gebruik gemaakt van ACT-R's geheugen, maar ook van ACT-R's visuele en motorieke systeem. Wat wij specifiek hebben toegevoegd zijn de eigenschappen van de werkgeheugen-module, en hoe die een rol spelen in multitaskinggedrag en -problemen.

Modelgebaseerde Neurowetenschap

Tot nu toe heb ik het alleen over gedrag (en pupilgrootte) gehad, maar dit gedrag komt natuurlijk voort uit onze hersenen. In Hoofdstuk 4 en 5 van dit proefschrift hebben we daarom gekeken naar waar de *problem state bottleneck* zich in het brein bevindt. Ook hierbij hebben we gebruik gemaakt van cognitieve modellen. In plaats van te kijken naar waar de hersenen actief zijn wanneer onze proefpersonen een experiment doen, zoals gebruikelijk is in de neurowetenschappen, hebben we in Hoofdstuk 4 gekeken naar hoe goed ons model hersenactivatie kan voorspellen. Op deze manier wilden we in meer detail beoordelen hoe goed of slecht het model is – als het naast reactietijden en fouten ook nog hersenactiviteit kan voorspellen, dan is het natuurlijk waarschijnlijker dat het een goede afspiegeling van de werkelijkheid is dan wanneer het 'alleen maar' reactietijden kan voorspellen.

In Hoofdstuk 5 hebben we daarnaast een nieuwe hersenanalysemethode toegepast: modelgebaseerde fMRI-analyse (fMRI is een techniek die laat zien welke delen van de hersenen actief zijn op een bepaald moment). Terwijl bij een standaard analyse gekeken wordt naar waar in het brein een experiment activatie veroorzaakt, wordt bij modelgebaseerde fMRI-analyse gekeken naar waar in het brein de onderdelen van een model zich bevinden. Dit klinkt ingewikkelder dan het is: de analyse vergelijkt simpelweg wanneer de onderdelen van een model actief zijn en wanneer de verschillende gebieden in de hersenen actief zijn. Als een specifiek onderdeel van een model en een hersengebied vaak tegelijkertijd actief zijn (vaker dan bij toeval het geval zou zijn), dan kan er geconcludeerd worden dat het betreffende hersengebied wellicht dit onderdeel van het model implementeert. Hierdoor is het mogelijk om heel precies de verschillende onderdelen van een model aan hersengebieden koppelen, preciezer dan over het algemeen mogelijk is met standaard analysemethoden. Wij hebben deze methode voor het eerst toegepast op een model dat geïmplementeerd is in een cognitieve architectuur. In de volgende sectie zullen we de resultaten hiervan bespreken, nadat we de resultaten van de gedragsexperimenten en het modelleerwerk besproken hebben.

Resultaten

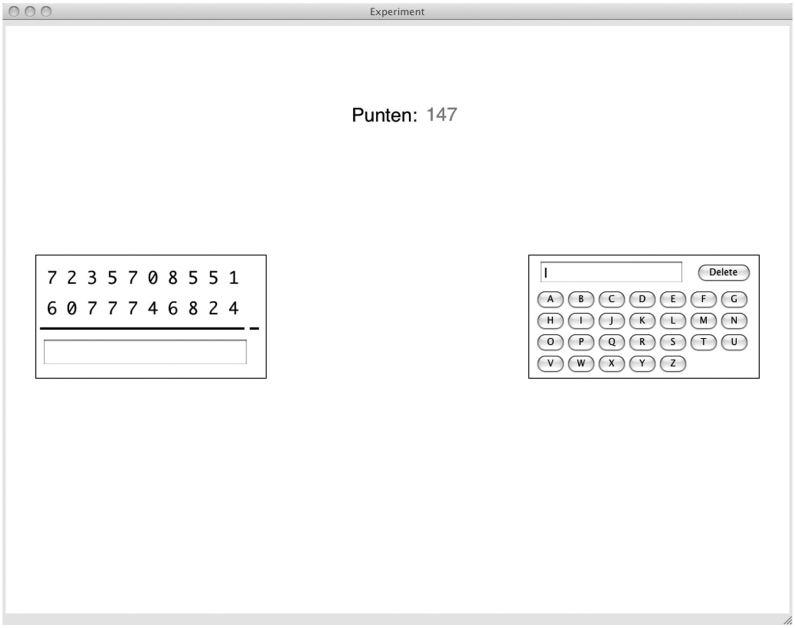
Gedragsexperimenten en Model

De Eerste Experimenten

Het onderzoek in dit proefschrift begon met de gedragsexperimenten in Hoofdstuk 2. Naar aanleiding van een voorspelling die voortkwam uit de multitaskingtheorie *threaded cognition* van Dario Salvucci en Niels Taatgen, hebben we onderzocht of het tegelijkertijd bijhouden van meerdere tussenresultaten tot een verslechterde multitaskingprestatie leidt. Om dit te onderzoeken hebben we onze proefpersonen twee taken laten uitvoeren: het maken van kolomaftreksommen en het invoeren van tekst (Figuur 8.1). Beide taken hadden twee condities: een makkelijke conditie waarin de proefpersonen geen tussenresultaat nodig hadden om de taak uit te voeren, en een moeilijke conditie waarin ze wel een tussenresultaat nodig hadden.

Voor de kolomaftreksommen betekende dit dat er in de makkelijke conditie nooit geleend hoefde te worden, terwijl er in de moeilijke conditie in elke som van 10 kolommen 6 keer geleend moest worden. Figuur 8.1 laat de moeilijke conditie zien, voor de eerste kolom (1 – 4) moet er bijvoorbeeld direct al geleend worden. Het idee is dat de proefpersonen als tussenresultaat steeds moeten bijhouden of ze wel of niet geleend hebben in de voorgaande kolom. Voor het teksttaak, aan de rechterkant van Figuur 8.1, moesten de proefpersonen in de makkelijke taak steeds op de letter klikken die getoond werd in het venstertje boven de letter-knoppen (in dit geval een 'I'). Zodra ze op een letter klikten verscheen er een nieuwe letter. In de moeilijke variant werd er steeds een woord van 10 letters getoond dat de proefpersonen moesten invoeren. Zodra de proefpersonen op de eerste letter klikten, verdween het woord, en moest het verder uit het hoofd ingevoerd worden. In deze taak moesten de proefpersonen dus bijhouden welk woord ze aan het invoeren waren en bij welke letter ze waren (bijvoorbeeld 'fluisteren, 5de letter'). Het interessante aan dit experiment is dat de proefpersonen steeds na elke letter en elk cijfer tussen de taken moesten wisselen. Dit betekende dat als een taak moeilijk was, het tussenresultaat bijgehouden moest worden terwijl er een respons op de andere taak gegeven werd.

Het achterliggende idee van het experiment was dat proefpersonen alleen in de conditie wanneer beide taken moeilijk waren meer dan één tussenresultaat nodig hadden. Onze hypothese was dat het daarom in deze conditie veel slechter zou gaan dan in de andere condities. Dit is precies wat er uit het experiment kwam: terwijl proefpersonen natuurlijk trager reageerden en meer fouten maakten in de moeilijke condities, werden ze extra traag en maakten ze extra fouten wanneer beide taken moeilijk waren. Dus omdat de *andere* taak ook moeilijk was – omdat ze voor de andere taak ook een tussenresultaat bij moesten houden – werden de proefpersonen trager en gingen ze meer fouten maken (in beide taken).



Figuur 8.1 De interface van het experiment.

Hoe kunnen we dit verklaren? Zoals ik hierboven al kort beschreven heb gaan wij ervan uit dat mensen een werkgeheugen-module in hun hersenen hebben die maar één tussenresultaat tegelijkertijd kan bevatten. Zodra er een tweede tussenresultaat bij komt wordt het eerste tussenresultaat automatisch naar het normale geheugen verplaatst. Informatie ophalen uit het normale geheugen kost tijd en kan fout gaan, wat logischerwijs tot tragere reacties en fouten kan leiden. In de context van het experiment betekent dit dat proefpersonen zonder problemen de taken konden uitvoeren zolang geen of één van de twee taken moeilijk was: dan was er namelijk maar één tussenresultaat nodig (natuurlijk werden de proefpersonen al wel langzamer in de moeilijke taken omdat bijvoorbeeld de aftreksommen moeilijker werden). Alleen wanneer beide taken moeilijk waren moesten proefpersonen steeds tussenresultaten opslaan en weer ophalen uit hun geheugen, wat leidde tot de hogere reactietijden en meer fouten in deze conditie.

Dit idee hebben we geïmplementeerd als een cognitief model, dat vervolgens precies dezelfde taak moest uitvoeren als de proefpersonen. De specifieke reactietijden en fouten van het model hebben we vergeleken met de resultaten van de proefpersonen, en het bleek dat het model het patroon van de proefpersonen bijna perfect kon verklaren (zie ook de figuren in Hoofdstuk 2). Onze conclusie is daarom dat een werkgeheugen-module die maar één enkel tussenresultaat kan vasthouden een goede verklaring geeft voor de resultaten van het experiment.

Hoewel ons model liet zien dat dit een mogelijke verklaring was, zijn er natuurlijk ook andere manieren om de resultaten van het experiment uit te leggen. In twee andere experimenten in Hoofdstuk 2 hebben we twee andere verklaringen getest. In het eerste experiment hebben we getest of de effecten misschien toe te schrijven waren aan het feit dat het vasthouden van informatie over het algemeen alle cognitieve processen vertraagt. Dus, in plaats van een werkgeheugen-module die maar één tussenresultaat kan vasthouden, kunnen er bij deze verklaring wel meerdere tussenresultaten worden bewaard, maar leidt dit tot een algehele vertraging van ons cognitieve systeem. Om dit te testen hebben we een vergelijkbaar experiment gedaan als het eerste experiment, maar nu moesten de proefpersonen steeds twee letters en twee cijfers achter elkaar invoeren voordat ze tussen de taken wisselden. Volgens onze verklaring zou er alleen een vertraging moeten optreden bij het eerste cijfer en de eerste letter, omdat dan de inhoud van de werkgeheugen-module gewisseld moest worden. Volgens de alternatieve verklaring zou de vertraging juist op beide stappen moeten optreden, omdat het cognitieve systeem algeheel vertraagd is. De resultaten waren eenduidig en in lijn met ons model: we zagen alleen een vertraging op het eerste cijfer en de eerste letter in een reeks van twee. In het derde experiment in Hoofdstuk 2 hebben we nog een tweede alternatieve verklaring getest – een beperking in ons taalsysteem – maar ook die verklaring bleek niet op te gaan.

Op basis van deze eerste set van drie experimenten concludeerden we daarom dat de meest plausibele verklaring voor onze resultaten een werkgeheugen-module is die maar één tussenresultaat tegelijkertijd kan vasthouden: de *problem state bottleneck*.

Pupilgrootte

In Hoofdstuk 3 van dit proefschrift hebben we het bewijs voor een *problem state bottleneck* verder uitgebreid met nog een gedragsexperiment. In dit experiment moesten de proefpersonen wederom kolomaftreksommen oplossen en tekst invoeren, maar nu werd er op het scherm getoond of de proefpersonen net geleend hadden in één van de condities. Dit betekende dat in deze conditie het tussenresultaat op het scherm stond, en proefpersonen het dus niet hoefden te onthouden. Zoals verwacht presteerden de proefpersonen in deze conditie beter dan wanneer het tussenresultaat niet op het scherm stond.

Tijdens dit experiment hebben we ook de pupilgrootte van de proefpersonen gemeten, om te kijken of de conditie waarin beide taken moeilijk waren ook samenging met een toename in mentale belasting. Het bleek dat de pupilgrootte ten opzichte van de grootte vóór het experiment inderdaad het meest toenam in de conditie waarin beide taken moeilijk waren. Net als met de reactietijden en fouten, nam ook de pupilgrootte onevenredig veel toe in vergelijking tot de condities waarin maar één van de taken moeilijk was. Dit wijst erop dat de mentale belasting veruit het hoogst was als – volgens ons model – de tussenresultaten in de werkgeheugen-module verwisseld moesten worden.

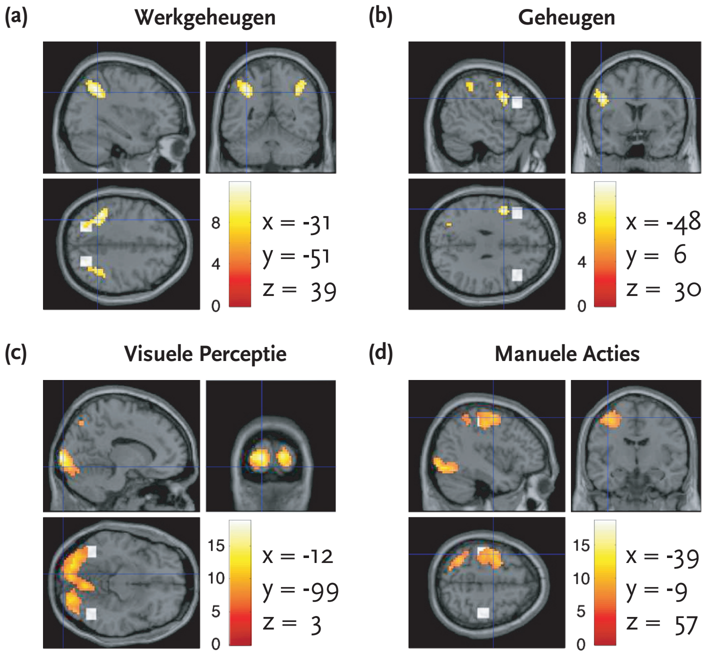
Uitwisseling via het Geheugen

Volgens ons model worden tussenresultaten die niet in de werkgeheugen-module passen opgeslagen in het normale geheugen. Volgens de gangbare geheugentheorieën in de psychologie verliest informatie in ons geheugen langzaam aan activiteit, en wordt het daarom steeds moeilijker om informatie op te halen die lang niet gebruikt is. Dat betekent voor onze theorie dat tussenresultaten die in het normale geheugen bewaard moeten worden langzamerhand steeds moeilijker zijn op te halen. We hebben dat getest in Hoofdstuk 6 met twee zogenaamde interruptie-experimenten. Interruptie-experimenten zijn bedacht om te kijken hoe mensen reageren op onderbrekingen van een taak. Het meest gebruikte voorbeeld hierbij is waarschijnlijk het schrijven van een lange email dat onderbroken wordt door een telefoontje. Na het beantwoorden van de telefoon kost het altijd weer even tijd om terug in het verhaal te komen. Interruptie-experimenten worden gebruikt om uit te vinden waarom dit het geval is.

Wij hebben twee experimenten gedaan waarin we respectievelijk de teksttaak en het maken van kolomaftreksommen onderbraken met een geheugentaak. Ook in deze experimenten hadden de taken weer een makkelijke en een moeilijke versie. De resultaten hiervan kwamen overeen met de andere experimenten: de proefpersonen waren het traagst en maakten de meeste fouten wanneer beide taken moeilijk waren. Interessanter echter was dat in dit experiment de onderbreking met de geheugentaak 4, 8, of 12 seconden kon duren. Geheugentheorieën in aanmerking genomen zou dat moeten betekenen dat proefpersonen de meeste moeite hadden om weer verder te gaan met de eerste taak na een onderbreking van 12 seconden, omdat dan de activiteit van het tussenresultaat in het normale geheugen het meeste was weggezakt. Volgens het model zou dit echter alleen het geval moeten zijn in de conditie waarin beide taken moeilijk waren, omdat alleen dan een tussenresultaat in het normale geheugen terechtkwam. De resultaten van de experimenten bevestigden deze voorspelling. Dit betekent dat tussenresultaten die niet in de werkgeheugen-module passen in ons normale geheugen worden opgeslagen. Het verklaart ook deels waarom interrupties van een taak vaak zo vervelend zijn. Neem het voorbeeld van het schrijven van een email dat onderbroken wordt door een telefoongesprek: na een telefoontje is de schrijver waarschijnlijk weer vergeten wat hij of zij op dat moment aan het schrijven was – het tussenresultaat – en moet dat weer opgehaald worden uit het geheugen (of opnieuw bedacht worden).

Neuroresultaten

Omdat de gedragsexperimenten en ons model op een *problem state bottleneck* wezen, hebben we in Hoofdstuk 4 en 5 geprobeerd om de locatie hiervan in de hersenen te bepalen. In Hoofdstuk 4 hebben we gekeken hoe goed ons model voorspellingen kan maken van hersenactiviteit in een aantal vooraf gedefinieerde gebieden. Over het algemeen bleek dit verrassend goed te zijn, wat aangaf dat het model een plausibele afspiegeling is van wat er in ons brein gebeurt als we dit soort taken uitvoeren. Omdat niet alle onderdelen van het model goede voorspellingen maakten, hebben we daarna in Hoofdstuk 5 gekeken wat de meest waarschijnlijk gebieden zijn om aan de verschillende onderdelen van het model te koppelen. Hiervoor hebben we de modelgebaseerde fMRI-analyse techniek gebruikt die ik eerder kort heb geïntroduceerd.

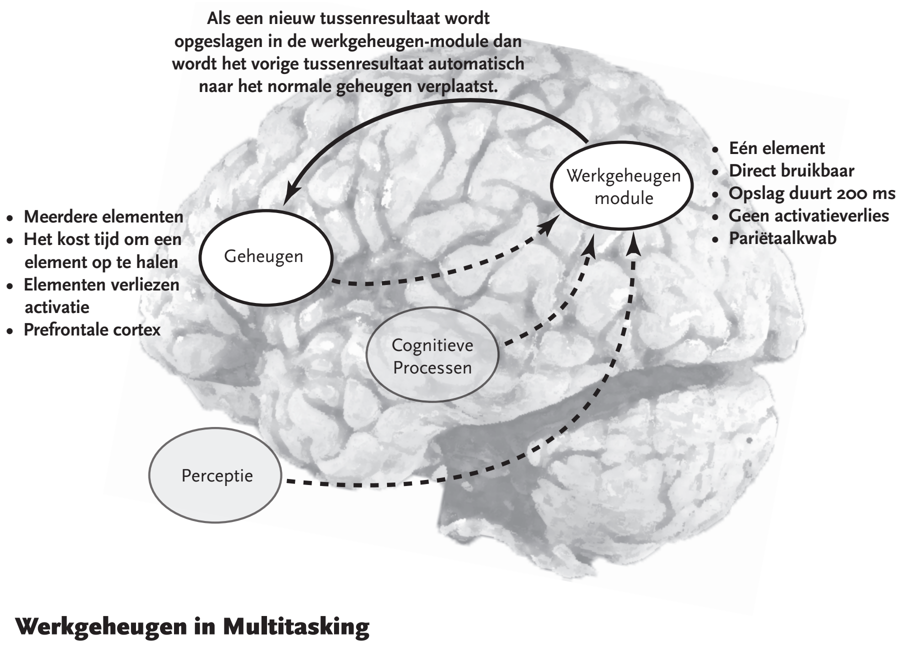


Figuur 8.2 Resultaten van de modelgebaseerde fMRI-analyse. De witte vierkantjes geven aan wat de vooraf gedefinieerde gebieden van de analyse in Hoofdstuk 4 zijn.

Modelgebaseerde fMRI-analyse laat zien wat de meest waarschijnlijke locatie van de verschillende onderdelen van een model is. In Figuur 8.2 staan de resultaten van de analyse: (a) laat de meest waarschijnlijke locatie van de werkgeheugen-module zien, (b) van het geheugen, (c) van de visuele perceptie – de ogen – van het model, en (d) van de manuele acties, dus van het gebruik van de muis en het toetsenbord van het model. Zoals verwacht bevond het meest waarschijnlijke gebied voor de visuele perceptie van het model zich helemaal aan de achterkant van ons brein, dit is namelijk het gebied dat standaard geassocieerd wordt met visuele perceptie. Ook de acties van het model werden gelokaliseerd op de verwachte locatie: de motor cortex, het gebied in de hersenen dat al onze bewegingen aanstuurt. De interessantere onderdelen van het model, de werkgeheugen-module en het geheugen bevonden zich volgens de analyse respectievelijk in de pariëtaalkwab en de prefrontale cortex. Ook dit is in overeenstemming met de literatuur.

Conclusie: Werkgeheugen in Multitasking

Gebaseerd op alle experimenten in dit proefschrift hebben we in Hoofdstuk 6 onze uiteindelijke theorie gepresenteerd: Werkgeheugen in Multitasking. Figuur 8.3 laat een overzicht van deze theorie zien, met alle elementen die ik hierboven al kort



Figuur 8.3 De uiteindelijke theorie: Werkgeheugen in Multitasking.

besproken heb. Het belangrijkst element hierin is natuurlijk de werkgeheugen-module in de pariëtaalkwab, die informatie kan uitwisselen met het geheugen in de prefrontale cortex. Omdat de werkgeheugen-module maar één tussenresultaat tegelijkertijd kan opslaan, leidt het gebruik van meerdere tussenresultaten – wat vaak nodig is in een multitasking situatie – al snel tot behoorlijke multitaskinginterferentie. Dit is één van de redenen waarom multitasking niet altijd een goed idee is. In dit proefschrift hebben we daarvoor ondersteuning gevonden, niet alleen met gedragsexperimenten, maar ook met behulp van een cognitief model en neurowetenschappelijke experimenten.

Acknowledgements

In the course of my PhD project I have been (forced to go) to a number of different project management courses and PhD retreats. A recurring topic was having advisor problems – either because advisors did not understand the PhD student's problems, did not spend sufficient time on the project, or were otherwise not really involved. These were the moments that I could show off with my own advisors.

During the last four years I slowly realized how lucky I was with having you, Niels and Hedderik, as my advisors: you always had time for me and were always supportive. Furthermore, you provided me with opportunities beyond what is normally expected. To give only two examples: Hedderik managed to get funding for me to go to a conference in Italy when I was still a master student (making it easier to subsequently get funding for my PhD project), and Niels let me use his grant money to carry out an fMRI experiment while I knew hardly anything about fMRI. Sometimes my project seemed to progress unfairly smoothly compared to other projects, and that was definitely thanks to your supervision!

Second, I would like to thank my committee: Dario, Jeroen, and Ritske. Thanks for reading through my dissertation, and thanks for being willing to come to my defense in Groningen.

Leendert and Enkhbold, thank you for being my 'paranimfen'. Leendert, you are probably the one who taught me most about how science 'works'. I am not sure who distracted whom most, but it was great to have you as a roommate. Enkhbold, it was fun complaining about facility management together and interesting to see you make the transition from computer to cognitive science. Also thanks to my other great roommates: Chris, Jacolien (who was always working hard, until Leendert mentioned soccer), Ben, and Anja.

Then, the Cognitive Modeling group. Our meetings were (almost) always interesting, the discussions very useful, and the atmosphere was great. Not only did the feedback improve the scientific quality of my work, but also my presentation skills. I believe that you can safely say that *we* won several poster prizes. Thanks: Niels, Hedderik, Fokie, Leendert, Jacolien, Jennifer, Marieke, Katja, Ben, Enkhbold, Tadeusz, Stefan, Jelle, Menno, Trudy, Florian, Udo and everyone else who came to our meetings.

The Artificial Intelligence department was a great environment to work in. Thanks for the gezelligheid, the many discussions on water kettles, and even sometimes on science. Thank you all!

Part of the work in this dissertation was carried out at Carnegie Mellon University. Without the help of John, Andrea, Jon, Jenn, Debbie, and Scott, the fMRI experiments in this thesis would not have been a success – thanks! Ion, thank you for beating me in squash every single time, even when you were injured.

Two experiments in this thesis were conducted by students I supervised: Trudy and Max. Thanks for the great work you did!

Finally, doing my PhD brought me much more than a title. Two conferences that I will never forget are the ACT-R Workshop and CogSci in 2008. During those conferences I met you, Katja. Thanks for always just understanding me!

References

Albert, N. B., Weigelt, M., Hazeltine, E., & Ivry, R. B. (2007). Target selection during bimanual reaching to direct cues is unaffected by the perceptual similarity of the targets. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1107-1116. doi: 10.1037/0096-1523.33.5.1107

Alm, H., & Nilsson, L. (1994). Changes in driver behaviour as a function of handsfree mobile phones–a simulator study. *Accident Analysis and Prevention, 26*(4), 441-451.

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review, 115*, 602-639.

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

Altmann, E. M., & Trafton, J. G. (2007). Timecourse of recovery from task interruption: Data and a model. *Psychonomic Bulletin & Review, 14*(6), 1079-1084.

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science, 26*, 85-112.

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29*, 313-341.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced tower of Hanoi. *Journal of Cognitive Neuroscience, 17*(8), 1261-1274. doi: 10.1162/0898929055002427

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060. doi: 10.1037/0033-295X.111.4.1036

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*, 341-380.

Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in Cognitive Science, 12*(4), 136-143. doi: 10.1016/j.tics.2008.01.006

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*: Lawrence Erlbaum.

Anderson, J. R., & Qin, Y. (2008). Using brain imaging to extract the structure of complex events at the rational time band. *Journal of Cognitive Neuroscience, 20*(9), 1624-1636. doi: 10.1162/jocn.2008.20108

Anderson, J. R., Qin, Y., Jung, K. J., & Carter, C. S. (2007). Information-processing modules and their relative modality specificity. *Cognitive Psychology, 54*(3), 185-217. doi: 10.1016/j.cogpsych.2006.06.003

Anderson, J. R., Qin, Y., Sohn, M. H., Stenger, V. A., & Carter, C. S. (2003). An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin & Review, 10*, 241-261.

Anderson, J. R., Qin, Y., Stenger, V. A., & Carter, C. S. (2004). The relationship of three cortical regions to an information-processing model. *Journal of Cognitive Neuroscience, 16*(4), 637-653. doi: 10.1162/089892904323057353

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*(3), 221-256.

Anderson, J. R., & Schooler, L. (1991). Reflections of the environment in memory. *Psychological Science, 2*(6).

Anderson, J. R., Taatgen, N. A., & Byrne, M. D. (2005). Learning to achieve perfect timesharing: Architectural implications of Hazeltine, Teague, and Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31*(4), 749-761. doi: 10.1037/0096-1523.31.4.749

Aue, T., Lavelle, L. A., & Cacioppo, J. T. (2009). Great expectations: What can fMRI research tell us about psychological phenomena? *International Journal of Psychophysiology, 73*(1), 10-16. doi: 10.1016/j.ijpsycho.2008.12.017

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

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. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839. doi: 10.1038/nrn1201

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). New York: Academic Press.

Bandettini, P. A., & Ungerleider, L. G. (2001). From neuron to BOLD: New connections. *Nature Neuroscience, 4*(9), 864-866. doi: 10.1038/nn0901-864

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276-292.

Borst, J. P., & Taatgen, N. A. (2007). The costs of multitasking in threaded cognition. *Proceedings of the eighth international conference on cognitive modelling* (pp. 133-138). Ann Arbor, MI.

Borst, J. P., Taatgen, N. A., Stocco, A., & Van Rijn, H. (2010). The neural correlates of problem states: Testing fMRI predictions of a computational model of multitasking. *PLoS ONE, 5*(9), e12966.

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*(2), 363-382. doi: 10.1037/a0018106

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2011). Using a symbolic process model as input for model-based fMRI analysis: Locating the neural correlates of problem state replacements. *NeuroImage, 58*(1), 137–147. doi: 10.1016/j.neuroimage.2011.05.084

Borst, J. P., Taatgen, N. A., Van Rijn, H., Stocco, A., & Fincham, J. M. (2009). Testing fMRI predictions of a dual-task interference model. *Proceedings of the 9th international conference on cognitive modeling* (pp. 24-30).

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage, 16*(2).

Broadbent, D. E. (1958). *Perception and communication*: Pergamon Press New York.
Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23*(4), 309-316.
Büchel, C., Wise, R. J., Mummery, C. J., Poline, J. B., & Friston, K. J. (1996). Nonlinear regression in parametric activation studies. *NeuroImage, 4*(1), 60-66. doi: 10.1006/nimg.1996.0029
Buckner, R. L., Kelley, W. M., & Petersen, S. E. (1999). Frontal cortex contributes to human memory formation. *Nature Neuroscience, 2*(4), 311-314. doi: 10.1038/7221
Buwalda, T. A., Borst, J. P., Taatgen, N. A., & van Rijn, H. (2011). Evading a multitasking bottleneck: Presenting intermediate representations in the environment. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33st annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review, 108*(4), 847–869.
Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *NeuroImage, 16*(2), 317-330. doi: 10.1006/nimg.2002.1063
Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*(1), 1-47.
Carrier, L. M., Cheever, N. A., Rosen, L. D., Benitez, S., & Chang, J. (2009). Multitasking across generations: Multitasking choices and difficulty ratings in three generations of Americans. *Computers in Human Behavior, 25*(2), 483-489. doi: 10.1016/j.chb.2008.10.012
Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386*(6625), 604-608. doi: 10.1038/386604a0
Cohen, M. S. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage, 6*(2), 93-103. doi: 10.1006/nimg.1997.0278
Collette, F., Hogge, M., Salmon, E., & Van der Linden, M. (2006). Exploration of the neural substrates of executive functioning by functional neuroimaging. *Neuroscience, 139*(1), 209-221. doi: 10.1016/j.neuroscience.2005.05.035
Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience and Biobehavioral Reviews, 26*(2), 105-125.
Coltheart, M. (2004). Brain imaging, connectionism, and cognitive neuropsychology. *Cognitive Neuropsychology, 21*(1), 21-25. doi: 10.1080/02643290342000159
Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? *Cortex, 42*(3), 323-331.
Cooper, R. (2007). The role of falsification in the development of cognitive architectures: Insights from a Lakatosian analysis. *Cognitive Science, 31*, 509-533.
Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology, 17*(4), 297-338.
Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(01), 87-114.

Culham, J. C., & Kanwisher, N. G. (2001). Neuroimaging of cognitive functions in human parietal cortex. *Current Opinion in Neurobiology, 11*(2), 157-163.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25*(3), 315-353. doi: 10.1016/S0364-0213(01)00039-8

Dale, A. M., & Buckner, R. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5*(5), 329-340.

Daw, N. D., & Frank, M. J. (2009). Reinforcement learning and higher level cognition: Introduction to special issue. [Editorial]. *Cognition, 113*(3), 259-261.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876-879.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*(3-6), 487-506. doi: 10.1080/02643290244000239

Diedrichsen, J., Hazeltine, E., Kennerley, S., & Ivry, R. B. (2001). Moving to directly cued locations abolishes spatial interference during bimanual actions. *Psychological Science, 12*(6), 493-498.

Dux, P., Ivanoff, J., Asplund, C., & Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron, 52*(6), 1109-1120.

Fletcher, P. C., & Henson, R. N. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain, 124*(5), 849-881.

Fodor, J. (1999). Let your brain alone. *London Review of Books, 21*(19), 68-69. Retrieved from http://www.lrb.co.uk/v21/n19/jerry-fodor/diary

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science, 6*(2), 78-84.

Friston, K. J. (2003). Introduction: Experimental design and statistical parametric mapping. In R. S. Frackowiak, K. J. Friston, C. Frith, R. Dolan & C. J. Price (Eds.), *Human brain function* (pp. 599-633). San Diego, CA: Academic Press.

Friston, K. J. (2009). Modalities, modes, and models in functional neuroimaging. *Science, 326*(5951), 399-403. doi: 10.1126/science.1174521

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (Eds.). (2007). *Statistical parametric mapping. The analysis of functional brain images.*: Academic Press.

Garavan, H. (1998). Serial attention within working memory. *Memory and Cognition, 26*(2), 263-276.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology, 44*(3), 297-337. doi: 10.1006/cogp.2001.0770

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

Gläscher, J. P., & O'Doherty, J. P. (2010). Model-based approaches to neuroimaging: Combining reinforcement learning theory with fMRI data. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 501-510. doi: 10.1002/wcs.57

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage, 9*(4), 416-429.

González, V., & Mark, G. (2004). Constant, constant, multi-tasking craziness: Managing multiple working spheres *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 113-120).

Gray, W. D. (2008). Cognitive architectures: Choreographing the dance of mental operations with the task environments. *Human Factors, 50*(3), 497-505.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*, 359-382.

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*(3), 461-482. doi: 10.1037/0033-295X.113.3.461

Hagoort, P. (2008). Should psychology ignore the language of the brain? *Current Directions in Psychological Science, 17*(2), 96-101.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience, 26*(32), 8360-8367. doi: 10.1523/jneurosci.1010-06.2006

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage, 21*(4), 1639-1651. doi: 10.1016/j.neuroimage.2003.11.029

Harley, T. A. (2004). Does cognitive neuropsychology have a future? *Cognitive Neuropsychology, 21*(1), 3-16. doi: 10.1080/02643290342000131

Haruno, M., & Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Networks, 19*(8), 1242-1254. doi: 10.1016/j.neunet.2006.06.007

Hazeltine, E., Teague, D., & Ivry, R. B. (2002). Simultaneous dual-task performance reveals parallel response selection after practice. *Journal of Experimental Psychology: Human Perception and Performance, 28*(3), 527-545.

Hendriks, P., Van Rijn, D. H., & Valkenier, B. (2007). Learning to reason about speakers' alternatives in sentence comprehension: A computational account. *Lingua, 117*(11), 1879-1896.

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *The Quarterly Journal of Experimental Psychology: Section A, 58*(2), 193-233. doi: 10.1080/02724980443000502

Henson, R. (2006). What has (neuro)psychology told us about the mind (so far)? A reply to Coltheart (2006). *Cortex, 42*(3), 387-392. doi: 10.1016/S0010-9452(08)70365-4

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science, 143*(3611), 1190-1192. doi: 10.1126/science.143.3611.1190

Hodgetts, H. M., & Jones, D. M. (2006). Interruption of the Tower of London task: Support for a goal-activation approach. *Journal of Experimental Psychology General, 135*(1), 103-115. doi: 10.1037/0096-3445.135.1.103

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*(2), 174-196. doi: 10.1145/353485.353487

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review, 116*(4), 717-751. doi: 10.1037/a0017187

Huss, D., & Byrne, M. (2003). An ACT-R/PM model of the articulatory loop. *Proceedings of the Fifth International Conference on Cognitive Modeling*, 135-140.

Iqbal, S. T., Adamczyk, P., Zheng, X., & Bailey, B. P. (2005). Towards an index of opportunity: Understanding changes in mental workload during task execution. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 311-320).

Janssen, C. P., & Gray, W. D. (in press). When, what, and how much to reward in reinforcement learning based models of cognition. *Cognitive Science*.

Jolicœur, P., & Dell'Acqua, R. (1999). Attentional and structural constraints on visual encoding. *Psychological Research, 62*(2), 154-164.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*(1), 193-224. doi: 10.1146/annurev.psych.59.103006.093615

Jonides, J., Nee, D. E., & Berman, M. G. (2006). What has functional neuroimaging told us about the mind? So many examples, so little space. *Cortex, 42*(3), 414-417.

Just, M. A., Keller, T. A., & Cynkar, J. (2008). A decrease in brain activation associated with driving when listening to someone speak. *Brain Research, 1205*, 70-80.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(756), 1583-1585.

Keele, S. W. (1973). *Attention and human performance*. Pacific Palisades, CA: Goodyear.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12*, 391-438.

Kieras, D. E., Meyer, D. E., Ballas, J., & Lauber, E. J. (2000). Modern computational perspectives on executive mental processes and cognitive control: Where to from here? In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes*. (pp. 681-712). Cambridge, MA: MIT Press.

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biology, 4*(8), e233. doi: 10.1371/journal.pbio.0040233

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

Kirsh, D. (1995). The intelligent use of space. *Artificial intelligence, 73*, 31-68.

LaBar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: A functional MRI comparison within subjects. *NeuroImage, 10*(6), 695-704. doi: 10.1006/nimg.1999.0503

Lacadie, C. M., Fulbright, R. K., Rajeevan, N., Constable, R. T., & Papademetris, X. (2008). More accurate Talairach coordinates for neuroimaging using non-linear registration. *NeuroImage, 42*(2), 717-725. doi: 10.1016/j.neuroimage.2008.04.240

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft, 8*(1), 5-19.

Lebiere, C., Gonzalez, C., & Martin, M. (2007). Instance-based decision making model of repeated binary choice. In R. L. Lewis, T. A. Polk & J. E. Laird (Eds.), *Proceedings of the 8th international conference on cognitive modeling* (pp. 67-72). London, U.K.: Taylor & Francis.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375-419. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10*(10), 447-454.

Liu, Y., Feyen, R., & Tsimhoni, O. (2006). Queueing network-model human processor (QN-MHP): A computational architecture for multitask performance in human-machine systems. *ACM Transactions on Computer-Human Interaction (TOCHI), 13*(1), 37-70.

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance, 5*(2), 189-207.

Logothetis, N. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878. doi: 10.1038/nature06976

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Cognitive Systems Research, 1*, 99-118.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(3), 817-835.

McFarlane, D. C., & Latorella, K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction, 17*(1), 1-61.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*(1), 3-65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104*(4), 749-791.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of experimental psychology Applied, 14*(4), 299-313. doi: 10.1037/a0014402

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review, 11*(2), 296-301.

Myung, I. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*(1), 190-204. doi: 10.1006/jmps.1999.1283

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review, 86*(3), 214-255.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). San Diego, CA: Academic Press.

Newell, A. (1980). Physical symbol systems. *Cognitive Science, 4*(2), 135-183.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1-18). New York: Plenum.

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences, 1104*, 35-53. doi: 10.1196/annals.1390.022

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411-421.

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 51, pp. 45-100): Academic Press.

Oberauer, K., & Bialkova, S. (2009). Accessing information in working memory: Can the focus of attention grasp two elements at the same time? *Journal of Experimental Psychology: General, 138*(1), 64-87. doi: 10.1037/a0014738

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review, 115*(3), 544-576.

Page, M. P. A. (2006). What can't functional neuroimaging tell the cognitive psychologist? *Cortex, 42*(3), 428-443.

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.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220-244.

Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention* (pp. 155-189). Hove, England: Psychology Press.

Pitt, M., & Myung, I. (2002). When a good fit can be bad. *Trends in Cognitive Science, 6*(10), 421-425.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*(5695), 443-447.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*(2), 358-367.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1*, 201-220.

Salvucci, D. D. (2005). A multitasking general executive for compound continuous tasks. *Cognitive Science, 29*, 457-492.

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors, 48*(2), 362-380.

Salvucci, D. D., & Bogunovich, P. (2010). Multitasking and monotasking: The effects of mental workload on deferred task interruptions *Proceedings of the 28th international conference on human factors in computing systems* (pp. 85-88).

Salvucci, D. D., & Macuga, K. L. (2002). Predicting the effects of cellular-phone dialing on driver performance. *Cognitive Systems Research, 3*, 95-102.

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

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review, 115*(1), 101-130. doi: 10.1037/0033-295X.115.1.101

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

Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. *Human factors in computing systems: CHI 2009 conference proceedings* (pp. 1819-1828). New York: ACM Press.

Schumacher, E. H., Lauber, E. J., Glass, J. M., Zurbriggen, E. I., Gmeindl, L., Kieras, D. E., & Meyer, D. E. (1999). Concurrent response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology: Human Perception and Performance, 25*(3), 791-814.

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., & Meyer, D. E. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science, 12*(2), 101-108.

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Sciences USA, 95*(3), 876-882.

Sohn, M. H., & Anderson, J. R. (2001). Task preparation and task repetition: Two-component model of task switching. *Journal of Experimental Psychology General, 130*(4), 764-778.

Sohn, M. H., Goode, A., Stenger, V. A., Jung, K. J., Carter, C. S., & Anderson, J. R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *NeuroImage, 25*(1), 21-33. doi: 10.1016/j.neuroimage.2004.11.001

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4*, 215-230.

Steinhauer, S. R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. *Annals of the New York Academy of Sciences, 658*, 182-204.

Stocco, A., & Anderson, J. R. (2008). Endogenous control and task representation: An fMRI study in algebraic problem-solving. *Journal of Cognitive Neuroscience, 20*(7), 1300-1314. doi: 10.1162/jocn.2008.20089

Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review, 117*(2), 541-574. doi: 10.1037/a0019077

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: The MIT Press.

Taatgen, N. A. (2005). Modeling parallelization and flexibility improvements in skill acquisition: From dual tasks to complex dynamic skills. *Cognitive Science, 29*(3), 421-455.

Taatgen, N. A., & Anderson, J. R. (2008). Constraints in cognitive architectures. In R. Sun (Ed.), *Handbook of cognitive modeling* (pp. 170-185): Cambridge University Press.

Taatgen, N. A., Juvina, I., Schipper, M., Borst, J. P., & Martens, S. (2009). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology*. doi: 10.1016/j.cogpsych.2008.12.002

Taatgen, N. A., Van Rijn, D. H., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review, 114*(3), 577-598. doi: 10.1037/0033-295X.114.3.577

Taatgen, N. A., & van Rijn, H. (2011). Traces of times past: Representations of temporal intervals in memory. *Memory & Cognition*. doi: 10.3758/s13421-011-0113-0

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14*, 1-35.

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

Trafton, J. G., Bugajska, M., Fransen, B., & Ratwani, R. (2008). Integrating vision and audition within a cognitive architecture to track conversations *Proceedings of HRI '08* (pp. 201-208).

Van Maanen, L., Van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review, 16*(6), 987-999. doi: 10.3758/PBR.16.6.987

Van Maanen, L., Van Rijn, H., & Taatgen, N. A. (in press). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*.

Van Rij, J., Hendriks, P., Spenader, J., & Van Rijn, H. (2009). Modeling the selective effects of slowed-down speech in pronoun comprehension. *Proceedings of the 3rd conference on generative approaches to language acquisition North America (GALANA 3)*. Somerville, MA: Cascadilla Press.

Van Rij, J., Van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language, 37*(3), 731-766.

Van Rijn, H., & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica, 129*(3), 365-375. doi: 10.1016/j.actpsy.2008.09.002

Vazquez, A. L., & Noll, D. C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage, 7*(2), 108-118. doi: 10.1006/nimg.1997.0316

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3*(4), 255-274.

Wagner, A. D., Maril, A., Bjork, R. A., & Schacter, D. L. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. *NeuroImage, 14*(6), 1337-1347. doi: 10.1006/nimg.2001.0936

Welford, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance—a review and a theory. *British Journal of Psychology, 43*(2).

Wickens, C. D. (1984). Processing resources in attention, dual task performance, and work load assessment. In R. Parasuraman, J. Beatty & R. Davies (Eds.), *Varieties of attention* (pp. 63-102). New York: Wiley.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2 ed.). New York, NY: HarperCollins.

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

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science, 12*(3), 219-224.

Wunderlich, K., Rangel, A., & O'Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences USA, 106*(40), 17199-17204. doi: 10.1073/pnas.0901077106

Zijlstra, F. R. H., Roe, R. A., Leonora, A. B., & Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology, 72*, 163-185.

Publication List

Journal Articles

Van Rijn, H., Dalenberg, J., **Borst, J. P.**, & Sprenger, S. A. (under revision). Memory Strength Correlates Negatively with Pupil Dilation in a Delayed Response Paired Associate Task. *Psychophysiology.*

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2011). Using a Symbolic Process Model as input for Model-Based fMRI Analysis: Locating the Neural Correlates of Problem State Replacements. *NeuroImage 58*(1), 137-147. doi: 10.1016/j.neuroimage.2011.05.084

Borst, J. P., Taatgen, N. A., Stocco, A., & Van Rijn, H. (2010). The Neural Correlates of Problem States: Testing fMRI Predictions of a Computational Model of Multitasking. *PLoS ONE, 5*(9): e12966. doi: 10.1371/journal.pone.0012966

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The Problem State: A Cognitive Bottleneck in Multitasking. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*(2), 363-382. doi: 10.1037/a0018106

Van Maanen, L., Van Rijn, H., & **Borst, J. P.** (2009). Stroop and Picture-Word Interference are Two Sides of the Same Coin. *Psychonomic Bulletin & Review, 16*(6), 987-999. doi: 10.3758/PBR.16.6.987

Taatgen, N. A., Juvina, I., Schipper, M., **Borst, J. P.**, & Martens, S. (2009). Too Much Control can Hurt: A Threaded Cognition Model of the Attentional Blink. *Cognitive Psychology, 59*(1), 1-29. doi: 10.1016/j.cogpsych.2008.12.002

Martens, S., Johnson, A., Bolle, M. & **Borst, J. P.** (2009). A Quick Visual Mind Can Be a Slow Auditory Mind: Individual Differences in Attentional Selection Across Modalities. *Experimental Psychology, 56*(1), 33-40. doi: 10.1027/1618-3169.56.1.33

Book Chapter

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (in preparation). Using Cognitive Architectures to Analyze fMRI Data of Complex Tasks. In A. Johnson & R. W. Proctor (Eds.), *Neuroergonomics: Cognitive neuroscience approaches to human factors and ergonomics*: Palgrave Macmillan.

Peer-Reviewed Conference Papers

Buwalda, T. A., **Borst, J. P.**, Taatgen, N. A., & Van Rijn, H. (2011). Evading a Multitasking Bottleneck: Presenting Intermediate Representations in the Environment. In *Proceedings of the 33st Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2009). Modeling Triple-Tasking without Customized Cognitive Control. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Borst, J. P., Taatgen, N. A., Van Rijn, H., Stocco, A., & Fincham, J. M. (2009). Testing fMRI Predictions of a Dual-Task Interference Model. In *Proceedings of the 9th International Conference on Cognitive Modeling*. Manchester, UK.

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 *Human Factors in Computing Systems: CHI 2009 Conference Proceedings* (pp. 1819-1828). New York: ACM Press.

Borst, J. P. & Taatgen, N. A. (2007). The Costs of Multitasking in Threaded Cognition. In *Proceedings of the Eighth International Conference on Cognitive Modeling*. (pp.133-138). Ann Arbor, MI.

Abstracts

Borst, J. P., Taatgen, N. A., Van Rijn, H. (2011). Using a Model-Based fMRI Analysis Method to Locate the Neural Correlates of a Multitasking Bottleneck. In *Proceedings of the 33st Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Borst, J. P., Taatgen, N. A., Van Rijn, H. (2011). Locating the neural correlates of a multitasking bottleneck: Analyzing fMRI data on the basis of a computational cognitive model. *Interdisciplinary Workshop on Cognitive Neuroscience, Educational Research and Cognitive Modeling*. Delmenhorst, Germany.

Borst, J. P., Taatgen, N. A., Van Rijn, H. (2010). Locating the Neural Correlates of the Problem State Resource: Analyzing fMRI Data on the Basis of a Computational Model. In *Proceedings of the 10th International Conference on Cognitive Modeling*. Philadelphia.

Borst, J. P. (2010). Using Cognitive Models to Analyze fMRI Data. *Seventeenth annual ACT-R workshop*. Philadelphia.

Borst, J. P., Taatgen, N. A., Stocco, A., & Van Rijn, H. (2008). Locating the Problem Representation Bottleneck in the Brain. *Fifteenth annual ACT-R workshop*.

Borst, J. P., Taatgen, N. A., Stocco, A., & Van Rijn, H. (2008). Problem Representations in Multitasking: An Additional Cognitive Bottleneck. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*.

Van Maanen, L., **Borst, J. P.**, Janssen, C. P. & Van Rijn, H. (2006). Memory structures as user models. In *Proceedings of the thirteenth annual ACT-R workshop*.

Borst, J. P. & Van Rijn, H. (2006). Memory Decay Problems in a Level-Repetition Switch Task Model. In *Proceedings of the Seventh International Conference on Cognitive Modeling*. (pp. 351-352). Triest, Italy.