Capacity Limitations in Human Cognition

Behavioural and Biological Contributions

Von der Philosophisch-humanwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Walter J. Perrig und Prof. Dr. Rudolf Groner
angenommen.

Bern, den 11.11.2005   Der Dekan: Prof. Dr. Norbert Semmer

ACKNOWLEDGEMENTS

CONTENTS

1. Introduction

The human brain is a most unusual instrument of elegant and as yet unknown capacity.

Stuart L. Seaton (1958)

Everything has its limit - iron ore cannot be educated into gold.

Mark Twain (1835 – 1910)

If you always put limit on everything you do, physical or anything else, it will spread into your work and into your life. There are no limits. There are only plateaus, and you must not stay there, you must go beyond them.

Bruce Lee (1940-1973)

In order to be able to set a limit to thought, we should have to find both sides of the limit thinkable (i.e., we should have to be able to think what cannot be thought).

Ludwig Wittgenstein (1921)

Are there capacity limitations in human information processing? If there are capacity limits, what are the behavioural and neuronal consequences of these limitations, if we reach and exceed those capacity limits? Are there interindividual differences in capacity, and if there are differences, what is their basis? Are capacity constraints fixed for every person, or is capacity extendable under certain circumstances? What would be the consequences in practical and theoretical domains, if capacity could be extended?

These questions have been the focus of my interest for some time now, and as the citations above show, there are many controversies in the literature (not only in the domain of psychology), whether there are capacity limitations, where to find them, and of what nature they are. In everyday life, capacity limitations manifest themselves quite often and they are present in everyone at every stage of development, with the most obvious phenomenon being forgetting. However, there

seem to be some individuals with a large processing capacity and there are many speculations, on how such performance can be achieved. For example, there are some spectacular but rather anecdotic stories in the literature about people with great attentional or memory capacity. One of the oldest examples stems from the Scottish philosopher Dugald Steward, describing that he was able to play the cembalo while speaking to another person without interference (Stewart, 1792; as cited in Neumann, 1992). Allport, Antonis, and Reynolds (1972) describe more of these phenomena and also succeeded to demonstrate such effects experimentally. It seems to be common sense that there are some tasks, which are easily combined without interference, such as taking notes while listening to a lecture, while others are very difficult to combine, such as talking to somebody while reading a textbook or the above mentioned example, which is certainly impossible for most of us. Thus, there are some persons, for whom the simultaneous performance of tasks is achieved quite easily and efficiently, while others have more trouble. Especially in brain-injured patients, one of the most commonly claimed problems is that they have difficulties in attending to more than one thing at the same time, a severe limiting factor in work and daily life (Zimmermann & Leclercq, 2002).

Thus, to be able to perform two or more tasks at the same time demands some sort of organization or rather, resource allocation, which in turn seems to be dependent on a certain amount of capacity. What is the nature of this capacity? Is it fixed and the same in every person? What are the mediating factors which have an effect on capacity? How are these capacity limits expressed in the brain on a functional level?

In the first part of my thesis, I will focus on the theoretical basis of capacity limitations, the historic development and early examinations of the concept in the domain of cognitive psychology, followed by some actual considerations, also in the domain of neuronal correlates of capacity limitations as assessed with functional magnetic resonance imaging (fMRI). In the second part, I will report experiments, which I conducted in order to shed some light on the questions formulated above. Some of these experiments were carried out by means of fMRI and some on behavioural basis:

In Experiment 1, I will describe the task (i.e., the n-back task) which was used in all following experiments, describing its properties, its reliability and validity, as well as its relation to other measures corresponding to the same cognitive domain.

In the next section, I will discuss the neural processes underlying capacity limits, and further, the mechanisms taking place in the brain if capacity limits are exceeded, investigating the hypothesis by D'Esposito (2001). He predicted that if capacity limits will be reached, a decrease in activation would be observed on neuronal basis, what was investigated with a study already published in 2003 (Jaeggi et al., 2003). In Experiment 2 and 3, this study was replicated with different material, mainly to extend its implication while comparing the activation patterns in different groups of participants according to their performance (Experiment 2), and further, to work with only high- and low-performing participants increasing the memory load to an extent where capacity limits are most certainly exceeded in every participant (Experiment 3). Experiment 4 compares age-related performance in these tasks, comparing a young and a middle-aged sample of participants, firstly on behavioural basis (Experiment 4a) and secondly, on functional basis (Experiment 4b).

Experiment 5 was conducted in order to explore the relationship of these capacity demanding tasks with intelligence measures, and in order to shed light on the basis of individual differences in these tasks.

Experiment 6 and 7 finally tackle the issue, whether capacity limitations are trait-like, i.e., fixed, or whether it is be possible to extend these limitations with training and whether generalized effects on other domains can be observed.

In the last section, all the findings are integrated and discussed, and further issues remaining to be investigated are pointed out.

2. The nature of capacity limitations – Theoretical issues

2.1. 'Capacity': Concept Definition

The word 'capacity' has its etymological roots in the 15th century and can be derived from the Middle English 'capacite'. It was introduced into the English language via the French word *capacité*, coming in turn from the Latin word *capacitas* (1. room, space; 2. ability, capability, fitness, suitability). Among other meanings, in Webster's Dictionary (1994) the term is referred to as 1. the power of receiving or containing (e.g.: *The hotel has a large capacity.*); 2. cubic content; volume; that which can be contained (e.g.: *The gasoline tank has a capacity of 20 galleons.*); 3. the power of receiving impressions, knowledge, etc.; mental ability, capability (e.g. *That book was written for those with the capacity of scholars.*); 4. actual of potential ability to perform, yield, or withstand (e.g.: *He has a capacity for hard work.*) (…).

In psychology, the term 'capacity' has been widely used in different domains and it has to be asked primarily: What do we mean with capacity? What is capacity about? And: Capacity of what?

Most usually, capacity is linked to the limits of something, e.g. of mental capacity, processing capacity, memory capacity, etc., and therefore implying some more or less fixed and limited amount of resources, being the basis and constraints of other cognitive tasks. Naturally, not all agree on that definition and to the domain, where capacity plays a part in cognition, and even less, where capacity limits come into play. Navon (1998) for example states that capacity has acquired a circumscribed sense in resource theory '– as a scientific term denoting the limit on resources' (p. 841). Therefore, capacity would be closely linked, if not be isomorphic, to some kind of resources and their availability as stated above. In the view of resource-theories, 'resources' refer to 'provisions for, or internal inputs to processing' (Gopher 1986; Kahneman 1973; Navon 1984; Navon & Gopher, 1979; 1980; Navon & Miller, 1998; Norman & Bobrow, 1975; Wickens 1984; as cited in Navon, 1998). Wickens (1984) however, preferred the term resources to the term capacity (and even to attention, or effort), because, *resources* reflect a 'variable commodity'. *Capacity* on the other hand, assumes a fixed amount, i.e., a 'maximum limit', whereas *effort* is connoted with motivational factors, and *attention* 'possesses a variety of ambiguous meanings' (Wickens, 1984, p. 67). Still, resources are also poorly defined, which has

consequences for research: For example, since the 'variable commodity' of resources is not properly defined, it is not easy to make predictions about performances either. I prefer to assume capacity as a fixed entity, which can be described, defined and measured. Variations in capacity can therefore be tackled with an interindividual-differences approach, predicting and explaining other cognitive abilities and performance failures. This notion is also shared in working memory (WM) research, e.g. by Oberauer, Süss, Schulze, Wilhelm, & Wittmann (2000).

To conclude, in psychology, capacity most commonly refers to some fixed maximum in the sense of mental resources, and there is no doubt about the existence of capacity limitations. Therefore, in the following chapters I will use the term as the maximum in mental performance, which can be achieved by an individual in a certain task. An open issue concerns the exact location of those limitations and what the mechanisms behind those are. Already defining and determining the maximum of capacity is a non-trivial issue, and maybe, the limits of this capacity are not as fixed as it may seem at the first thought.

2.2. Capacity limitations in cognition

In one of the most famous and influential classical works in psychology, 'The Principles of Psychology' (James, 1890), one chapter addresses the question on 'how many things we can attend at once', and there, James describes the properties of the 'span' of consciousness, or attention, respectively. According to James, there is no limit in the number of *things* we are able to attend to, rather, it depends '(…) on the power of the individual intellect, on the form of the apprehension, and on what the things are' (p. 405). And: 'When apprehended conceptually as a connected system, their number may be very large' (p. 405). In these two sentences, many issues are anticipated, which were addressed and experimentally tackled almost one hundred years later: The question of individual differences and intelligence (e.g. Engle, Kane, & Tuholski, 1999), the structural differences of the material (e.g. Wickens, 1980, 1984), and, the issue of chunking and binding (e.g. Miller, 1956; Treisman & Gelade, 1980). Later in the chapter, James also takes practice into consideration, that is, that usually, only one stimulus can be processed effortlessly and at once, unless '(…) *the processes are very habitual.*' (p. 409), a topic later extended by Shiffrin and Schneider's concept of automatic vs. controlled processing (Shiffrin & Schneider, 1977). James cautiously remains unspecific about the number of objects or stimuli, which can be processed at the same time, due to the various circumstances which play a part in processing as described above. However, Wundt (1896) proposes very specific numbers as to the maximal 'extent' of attention, i.e., an attention limitation to six simple 'impressions' ('Eindrücke') being very constant over various conditions. Yet, these six elements can be extended up to 30 elements by meaningful combinations, such as letter strings being easily remembered if they make up a sentence (Wundt, 1896). The assumption of a fixed number of elements, or 'information units', being possible to process and remembered at one time has been later adopted by Miller (1956) in his famous article 'the magical number 7 plus or minus 2', which is still very present and widely cited in working memory research (see below).

The idea, that at any given time, the human information processing system can only handle a limited amount of information without interference to make use of that information for behaviour was the basis of capacity research which was taken up in the fifties of the last century, and mainly concentrated on the domain of attention. It is

assumed, that the filtering of irrelevant information has to be accomplished via selective attention mechanisms. Different approaches tried to explain (attentional) limitations which can be observed at certain circumstances. Two main directions and fundamentally different ways to this issue can be distinguished: Structural theories on one end, and capacity, i.e., resource theories on the other.

In *structural theories*, the so-called 'bottleneck', constraining the amount of information being processed, was the main issue to investigate. The core of these approaches lies in the *structure*, i.e., in the assumption, that the human information processing system has certain limitations (the 'bottleneck'), defining the border of parallel processing, after which only a limited amount of information can be processed consciously and/or attentively.

The first attempts to explore such selective processes experimentally were carried out by Cherry (1953) and Broadbent (1958), both focusing on the auditory modality. These paradigms led to the first information-processing theory of attention and were the starting point for many theoretical debates and the development of additional models of attention. Broadbent's (1958) 'bottleneck' or filter theory states that if two stimuli or messages presented at the same time, both gain access to a sensory buffer in parallel. Only one of the inputs is then allowed through a filter on the basis of physical characteristics (e.g. intensity) with the other input remaining in the buffer and not being analyzed and therefore not reaching consciousness ('all-or-nothing' principle). This locus of selection takes place at a very early stage in information processing. The filter, which lets through only one input at a time, prevents overloading of the limited-capacity mechanism beyond the filter (central processor); this mechanism finally, processes the input thoroughly (e.g. in terms of its meaning).

However, there has been a controversy between early- and late-selection theorists as to the nature and the location of this bottleneck in processing. Later research for instance showed that Broadbent's filter theory was inconsistent with findings, such as that selection can be based on the meaning of presented information (Gray & Wedderburn, 1960) and not only on physical properties. Treisman's attenuation theory (Treisman, 1964) could account for more of the findings. In this theory, the filter reduces or attenuates the analysis of information outside focal attention. The location of the bottleneck and the extent of such processing is seen as flexible, being determined in part e.g. by task demands (Johnston & Heinz, 1978), therefore being able to react in

the face of danger, despite concentrating on different tasks. To summarize, structural theories are concerned with two different processes taking place in one 'structure': Early, more automatic and effortless processes versus more attentive, effortful processes, which take place somewhat later in the processing stage. The border between these two processes is occupied by the bottleneck, remaining a controversial issue between the different approaches.

In *capacity* or *resource theories*, one of the challenging questions concerning attention is, why and how we are able to process more than one stimulus at once, and why we fail to do so in certain circumstances. No structural bottleneck is assumed in these theories; rather, several and parallel processes are possible from the first perceptual analyses up to the higher cognitive decision processes. Still, the interest in these theories also lies in performance failures, which is mainly investigated with the dual-task paradigm. Theoretically, the breakdowns of performance when two tasks are carried out simultaneously show the limitations of the human information-processing system. Some theorists argue that such breakdowns reflect the limited capacity of a single multi-purpose central processor or executive, sometimes described as 'attention' (the 'central capacity theories'; e.g. Kahneman, 1973; Norman & Shallice, 1986; see below); whereas other theorists favour the notion of several specific and independent pools of processing resources or modules (the 'multiple capacity theories', e.g. Allport, 1989; Wickens, 1980, 1984). In proposing more than one resource pool however, the difference to the structural theories is not clear, since these resource pools could also stand for a structural property.

The most common administered method for resource theorists in order to challenge capacity limitations is the dual-task paradigm. According to Oberauer, Lange, and Engle (2004), capacity is most purely measured in dual tasks, whereas in single tasks, strategies and/or other sources of variance seem to be involved to a larger extent. Oberauer et al. (2004) state that if two tasks are combined to a dual task, the strategies usually applied under single-task conditions will most probably fail because of interference, and therefore resulting in a purer measurement of processing capacity. Still, to my view, this notion might apply only to a certain extent, because I think that also in dual-task situations strategies can be used and are most certainly applied by participants, especially when the task is practiced. But I agree that a) usually, dual

tasks are generally more difficult than single tasks, and b) that the use of strategies poses a considerable problem to the theory of capacity limitations.

In the following section, I will describe several factors, which limit dual-task performance and which are mostly discussed in the context of resource theories: Task similarity (Wickens, 1984), task difficulty (Kahneman, 1973), and practice (Shiffrin & Schneider, 1977).

Task similarity refers to the fact that tasks interfere, if they have the same stimulus modality (e.g. auditory or visual), make use of the same stages of processing (input, internal processing, output), and rely on related memory codes (e.g. auditory or visual) (Wickens, 1984). This has already been described in1887 by Paulhan, who experimented with different dual-task conditions, such as writing a sentence while speaking another unrelated sentence (Paulhahn, 1887; as cited in James, 1890). The nature of dual-task interference due to task similarity has also been extensively studied and experimentally demonstrated by Pashler (e.g. 1984; 1994).

Task difficulty is a limiting factor in performance, a pattern demonstrated experimentally for example by Sullivan (1976). In the view of the central capacity theory (Kahneman, 1973), this central processor can be used flexibly across a wide range of activities but has strictly limited resources and is described as 'attention' or 'effort'. In Kahneman's (1973) view, there is no interference if two tasks do not demand the whole capacity of this processor; however, if task difficulty increases, the processor will be overloaded at some point, and therefore, performance is expected to decrease. Some experimental findings seem to be in line with the predictions of the central capacity theory (e.g. Bourke, Duncan, & Nimmo-Smith, 1996), but there are some important drawbacks: The nature of the central capacity is not well specified and remains rather descriptive, as well as the definition of 'task difficulty' and 'effort' remains vague. Further, task similarity seems to be a more important factor than task difficulty explaining interferences in some tasks (see for example Segal & Fusella, 1970). Therefore, according to Allport (1989) the findings speak for a modular, rather than a unitary attention system. Still, among other difficulties, the co-ordination of the outputs of several parallel organized modules is difficult to explain without the notion of a central capacity and there is also evidence that interference may occur in dual-task situations, even if very different resources are being tapped (Bourke et al., 1996).

Practice is somehow related to task difficulty and aids to dual-task performance either by the adoption of new and more effective strategies to minimize interference, or by reducing the attention resources needed to perform the task in shifting the controlled or effortful into more automatic processing (e.g. Shiffrin & Schneider, 1977). Automatic in contrast to controlled processing mechanisms make minimal demands on attentional resources; it seems therefore reasonable to assume that the processing ability is larger in practiced abilities. However, it is important to note, that through automatization, not the capacity itself is extended (assuming capacity as fixed entity), but rather, the *efficiency* of processing. Therefore, with increasing automatization, less resources are needed and thus, the demands on the task and the capacity decreases. Still, it is not easy to define automatization, and little consensus is found between researchers. In the context of frequency processing, a detailed definition and implications of automatic processing have been proposed by Hasher and Zacks (Hasher & Zacks, 1984; Zacks & Hasher, 2002): A key characteristic of automatic processing is seen in the minimal demands placed on the attentional system. Further defining criteria are summarized as follows: They

'(…) (a) operate without intention; (b) do not benefit from intention to encode the particular attribute; (c) do not benefit from training to encode the attribute or from feedback; (d) show minimal individual differences; (e) show minimal age differences; and (f) show minimal impact of state (e.g. arousal level) and situational (e.g. divided attention demands) factors that otherwise impact on available attentional capacity.' (Zacks & Hasher, 2002; p. 25).

Expanding the view of Shiffrin and Schneider (Shiffrin & Schneider, 1977), which are mainly concerned with the processing of perceptual inputs, Norman and Shallice (1986) proposed a framework in order to understand willful and conscious control of actions. The framework was mainly developed in order to explain phenomena as 'the fact that action sequences that normally are performed automatically can be carried out under deliberate conscious control when desired, and the way that such deliberate control can be used both to suppress unwanted actions and to enhance wanted ones.' (Norman & Shallice, 1986; p. 3). In their theory, two complementary components for the selection and control of actions are proposed, i.e., Contention Scheduling (CS) and a Supervisory Attentional System (SAS), also referring to neuropsychological evidence (i.e., from patients with frontal lobe damage

showing markedly deficits in tasks assumed to demand attentional control). CS corresponds to the control of behaviour in routine situations, where action sequences, usually relying on known schemata and habits, are implicitly guided by cues provided by the environment. It is also responsible for resolving conflicts between competing schemata. In case of insufficient routine control, e.g. in new situations or when there is a need for attentional control in case of an important or dangerous action sequence, the attentionally limited SAS comes into play, providing the source of control upon the selection between various schemata. The SAS is not further specified, e.g. in terms of modality-specific constraints, rather, it is seen as all-purpose processor in the sense of Kahneman's central processor (Kahneman, 1973).

To conclude, resource or capacity theories are primarily concerned with the impact of different task-related effects on dual-task processing and on tasks demanding high attentional control, yet, the issue remains unresolved, whether one single capacity in the sense of the SAS, or different resource pools do best account for performance in dual-task situations. A further, unresolved topic concerns the possible processes taking place when two tasks are executed simultaneously: One possibility is, that people switch or shift between the different tasks (serial processing), whereas the other possibility is that they really divide their attention equally between the two tasks (parallel processing). Both possibilities are not easily proved and are of importance when we talk about dual-tasking. In the view of Duncan (1980) however, it does not matter, whether the serial or the parallel approach is correct: The important fact to be noted is that performance decreases in either condition: "Either one process must wait (and may lose relevant information) while the other takes place, or both processes occur together, but neither with sufficient allocated resource" (Duncan, 1980; p. 77).

The view that controlled attentional processes have considerable impact on higher-order cognitive processes has also been acknowledged and adopted in the domain of working memory (WM). WM is defined as the set of mental processes, i.e. storing and processing of limited information over a short time in the service of ongoing higher-order cognitive functions (e.g. Cowan et al., 2005). As multi-modal framework, the concept has been introduced by Baddeley and Hitch (1974). It consists of storage buffers for verbal and visuospatial material; hierarchically above these buffers, a central executive system (CES) is postulated, being responsible for the

transfer from and to these buffers, for the manipulation of the memoranda, and as interface between the actual inputs and long-term memory (LTM). The CES is closely linked to the concept of controlled attention (Norman & Shallice, 1986) as described above. Capacity and its limitations are explicitly defined as feature of the WM model, since the capacity of every component is assumed to be limited; however there is much controversy, whether the assumed capacity is the same for all components, or whether there are separate capacities limiting performance in each WM component independently. A few years ago, Cowan (2001) launched a discussion around the long assumed 'magical' number seven in 'immediate memory' (Miller, 1956), suggesting that seven chunks of information proposed by Miller (1956) are far too highly estimated. Rather, he states that the attentional system is limited to four chunks of information (Cowan, 2001). This capacity limitation is termed as 'focus of attention' and is not dependent on input/output modality, thus, proposing a modality-free working memory capacity (WMC) in the sense of the SAS. Yet, defining WMC on numerical basis seems to be a controversial issue remaining superficial in many aspects. Halford, Wilson, and Phillips (1998) as well as Heitz, Unsworth, and Engle (2005) stress that WMC cannot be accounted for with the definition of a number or an amount per se; rather, qualitative differences, such as efficiency and complexity, which always depend on strategies (see below) contribute to the observed individual differences in WMC.

In a confirmatory factor analysis, as well as with structural equation modeling Oberauer, Süss, Schulze, Wilhelm, and Wittman (2000) investigated whether WMC would be best explained as unitary or as consisting of multiple factors. Indeed, they could demonstrate a domain-specific distinction between verbal and spatial WM (content dimension), as well as a distinction between storage/transformation/ coordination, a factor labelled as supervisory control (functional dimension). Thus, WMC seems to be best differentiated theoretically along two dimensions, i.e., contents and functions, which is also supported by the neuroimaging literature (e.g. Fletcher & Henson, 2001). According to Miyake and Shah (1999), there is now enough evidence that there is no single, capacity-limiting factor in human information processing as described above (Kahneman, 1973; Norman & Shallice, 1986), but rather, that more than one capacity will come into play in different situations, depending on the tasks to be performed, and also depending on the amount of

domain-specific knowledge and skill available to an individual (Ericsson & Delaney, 1998).

The use of strategies and skills are challenging for capacity research, and usually, researchers try to prevent the possibility of using a strategy in order to estimate a 'pure' measure of WMC (e.g. Cowan et al., 2005; Oberauer et al., 2004). However, the use of strategies cannot be ruled out in most of the cases, suggesting to explicitly investigate the strategies applied by the participants: Asking participants about their strategies could also reveal important mechanisms underlying processing at capacity limitations, although the gain in this procedure remains clearly explorative, since strategies are not always (verbally) accessible. In dual-task situations for example, the commonly observed lower performance in one of the tasks could merely reflect the participants' trade-off behaviour, in paying more attention to one task only in order to save resources. Only relying on the behavioural data, it would have to be concluded, that one task might be easier and more automatically applied than the other, thus, inadequately interpreting the data.

Another participant-related issue contributing to observed capacity constraints concerns motivation. Motivational factors seem to be important in controlled processing, and Norman and Shallice (1986) argue that these factors add to the optimal functioning of the SAS. Motivation also seems to contribute to considerable variance in WM performance (see for example Pochon et al., 2002), and additionally, Visser and Merikle (1999) demonstrated that the motivational state of the participants has substantial impact on perceptual processes; however, in memory tasks, this influence was not observed. Concluding to this subject, it seems appropriate to mention some of Duncan's (1980) reflections, concerning the relationship between observed performance decrements and inferred capacity limitations: It is not always clear where capacity limits come from and how much performance decreases can be explained with capacity limits, or whether other confounding factors, i.e., 'unspecific' source of variance contribute to the decrements, such as the simple increase in the number of processes, or motivation, among many others. The main issue in capacity research should therefore concentrate on the question, whether performance decrements are specific and clearly attributable to capacity limitation, or if they are more of an 'unspecific' nature. Although, it seems obvious, that to some extent, both

contribute to the observed performance decreases and it will be difficult to separate the two adequately.

To end this chapter, I would like to add that it might be confusing to speak about capacity limitations in different domains, such as attention and WM. Nevertheless, the issues discussed above show that these concepts share many features, the most prominent factor being controlled processing. Still, there are various factors and/or constraints which contribute to capacity limitations, always depending on the focus of interest or the domain in which the research takes place. To me however, it seems not very important to the question of interest, whether we speak of 'attention', 'WM', or other constructs in order to explain capacity limitations; rather, as stated above, the most important issue would be to find and define the performance limits and to make predictions about other related or general cognitive functions.

3. How to measure capacity limitations – Methodological issues

3.1. General issues

This turns us to the issue how capacity limits can be measured. Operationally defined, a load manipulation can be undertaken by either parametrically varying the load placed on a single process (quantitative variation), or by changing the actual process that is being loaded (qualitative variation). In the former, most straightforward case, a task is held constant but the *quantitative* demands placed on processing associated with that task are varied. For example, in the typical working memory study, the task - actively holding information in memory - remains unchanged, but the amount of information held in memory is parametrically varied in increasing the quantity of the memoranda (for example by enlarging a memory list length). There are many examples for such tasks, while the most typical parametric task being certainly the Sternberg task (Sternberg, 1966), where participants have to memorize a series of items (e.g. digits) for a short time. After a short delay, a probe is presented, and the participants have to indicate whether or not the probe was part of the previously presented items. Response times and errors increase linearly with increasing amount of items in the memory set, therefore the task is very appealing in order to assess processing capacity in individuals. Other tasks in this domain are the Reading Span Task (RST) as introduced by Daneman and Carpenter (1983), various simple span tasks like the Digit- Word- or Corsi Block Span (Lezak, Howieson, & Loring, 2004), or the self-ordered pointing task (SOPT; Petrides & Milner, 1982). Processing demands can also be induced by *qualitatively* changing the involved processes, for instance by introducing a secondary task, which is to be performed simultaneously with the first task (dual-task paradigm).

These quantitative or qualitative process changes, which indicate underlying capacity limitations, manifest themselves behaviourally in performance decrease and should be predictable from theoretical assumptions about the nature of the capacity limitations (Oberauer, 2005; Oberauer & Kliegl, 2001). In terms of Sternberg (1966), underlying such performance decreases 'is the supposition that if the selection of a response requires the use of information that is in memory, the latency of the response will reveal something about the process by which the information is retrieved (p. 652)', i.e., something about the mechanisms underlying capacity limitations.

This turns us to the behavioural measures of performance which are assumed to be markers, i.e., *reaction times* (RTs) and *accuracy* (i.e., error rates) serving as dependent variables for capacity limitations. Increasing response latencies, i.e., RTs are commonly observed with increasing memory load (Sternberg, 1966). There are some very early discussions of capacity limits manifesting themselves in RTs: In the fifties, a logarithmic increase in RTs with the number of alternatives in a choice situation was observed and interpreted in terms of capacity limitations (e.g. Hick, 1952; Hyman, 1953; as cited by Duncan, 1980). Similarly, accuracy as measured in error rates is also considered as indicator for capacity limitations, i.e., with increasing demands, the error rates usually increase.

To define capacity limits with performance measures however is not a trivial question: At which point can we state that performance 'has suffered' enough to be indicative as capacity limitations? Concerning accuracy, some define the limits as the criterion, where performance is at chance. Broadbent (1975) noted that usually the span measure is defined as the number of items that can be recalled on 50% of the trials which applies for example to the RST: Working memory capacity (WMC) is defined as stage, where more than 50% of the to be remembered items are remembered correctly. Others like Callicott et al. (1999) argue that not performing at ceiling levels anymore, i.e., that any performance discontinuities indicate that limitations in processing capacity are reached. This assumption goes back to Miller, assuming that the 'span of immediate memory' (the number of items that can be recalled without error) and the 'momentary capacity of consciousness' (the number of items immediately available) is reflecting the capacity limitation (Miller, 1962; as cited in Sternberg, 1966; p. 652). Both approaches might be correct to some extent but make it difficult to define an exact border of the capacity limitations (see Figure 1), which is even more challenging, if RTs are taken as performance measure.
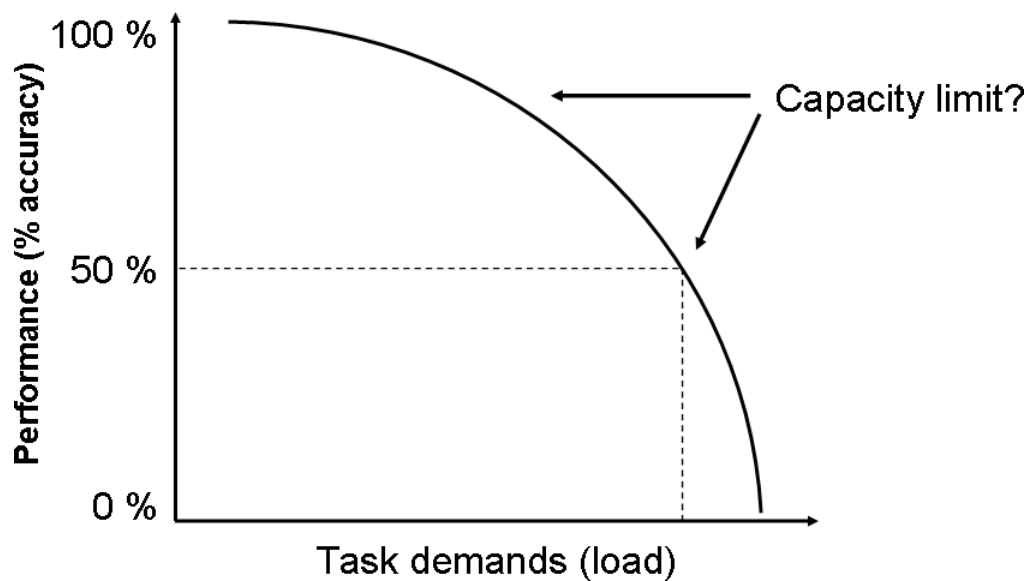
Figure 1. Hypothetical curve representing a performance decrease with increasing processing load, with possible locations of a capacity limitation.

Another issue raised by Navon (1998) addresses the question about the nature of capacity limitations reflected in performance decreases: Does it reflect the capacity of the output system (for example the manual responses) or a limit of resources? Also, this question is not easily answered.

Further, we have to take the participants' strategies into account, which might consider either to respond fast or to give more accurate responses (also depending on instructions given to the participant); in fact, usually a speed-accuracy trade-off in information processing is observed (Wickelgren, 1977). Therefore, it is important to investigate both behavioural measures, RTs *and* accuracy, in order to estimate their differential impact on the issue of capacity limitations.

3.2. The n-back task as a tool to measure capacity limitations

An alternative to the tasks described above in order to explore capacity and its limitations is the so-called parametric n-back task. It has attracted much interest by the recent neuroscience research, and it was originally introduced by Kirchner (1958) as visuospatial task with four load-factors ('no-back' to 'three-back'), and also by Mackworth (1959), who used up to 6 load-factors as visual letter task. Gevins et al. (1990) established this task in neuroscience in using it as 'visuomotor memory task' with one load-factor (2-back) for Event Related Potentials (ERP). To date, the n-back task is widely and almost uniquely used in various domains of functional neuroimaging (ERP, fMRI, and Positron Emission Tomography; PET). In the early days of fMRI-research, i.e., in the late nineties of the last century, the n-back task was growing popular because it was a task easy to apply but also one of the only tasks yielding reliable activation in prefrontal areas.[1]

The task consists of sequentially presented stimulus material (e.g. letters) and the participant is required to respond positively whenever the current stimulus matches the one presented *n* positions back in the sequence (see Figure 2). The value of *n* is commonly regarded as proportional to processing load, which is manifested in increasing error rates and RTs as *n* increases.

---

[1] E.E. Smith, personal communication, July 4th, 2004.
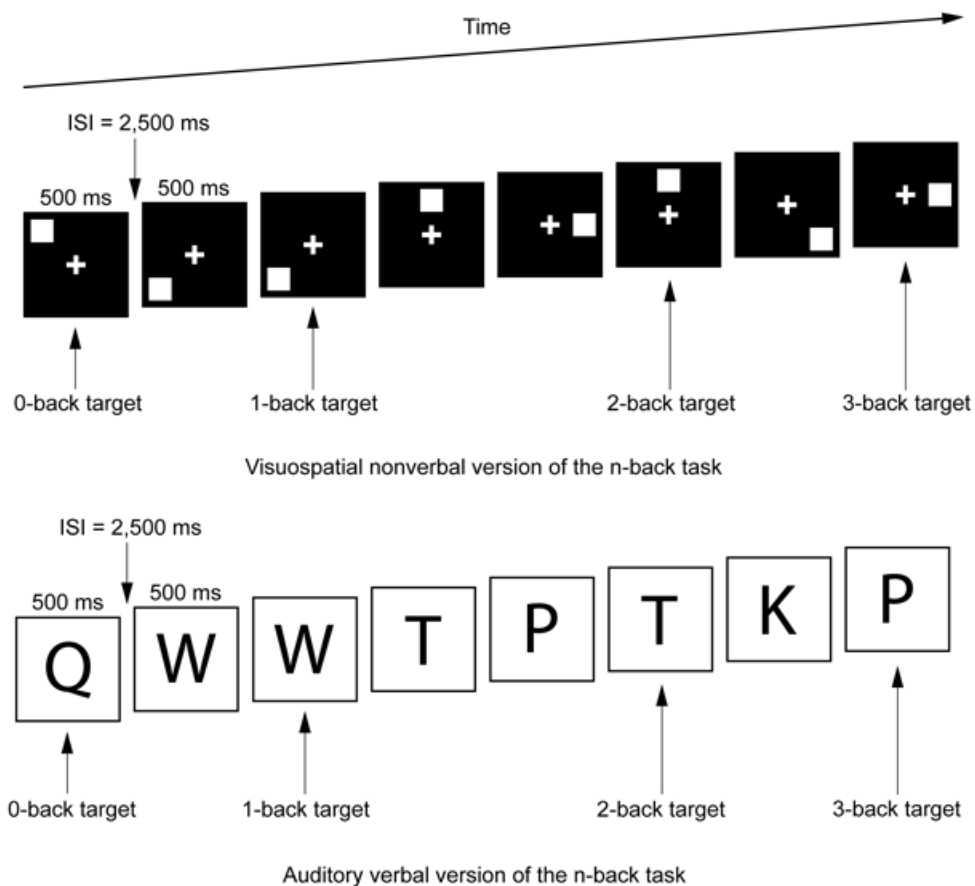
Figure 2: Four memory conditions for the sequential n-back task illustrated for a visuospatial and a verbal condition. Stimuli are presented sequentially, and for the 0-back condition, usually serving as baseline condition with minimal memory demands, the participant is requested to respond to a prespecified target (i.e., "Q" for the auditory condition, and 'upper left corner' for the visual condition), whereas for the other conditions, a response is required whenever the current stimulus matches the stimulus 1, 2, or 3 positions back in the sequence. ISI: Interstimulus Interval.

The attractive property of this task is the possibility to continuously manipulate processing load from a very easy level up to a level, which is usually quite challenging for some participants (3-back and more). Many kinds of different stimuli have been used so far: Letters (e.g. Braver *et al.*, 1997), words (e.g. Barch, Sheline, Csernansky, & Snyder, 2003), digits (e.g. Hoshi *et al.*, 2003), shapes (e.g. Jaeggi *et al.*, 2003), objects (e.g. Kubat-Silman, Dagenbach, & Absher, 2002), faces (e.g. Druzgal & D'Esposito, 2001), abstract patterns (e.g. Owen *et al.*, 1998), locations (e.g. Carlson *et al.*, 1998), pitch (e.g. Anourova *et al.*, 1999), colour (e.g. Vuontela, Rama, Raninen, Aronen, & Carlson, 1999), and many more. The behavioural data in all of these versions seem to be comparable, that is, irrespective of the material being used,

increasing RTs and error rates are observed with augmenting n. Since the task is used primarily in functional neuroimaging, a few words to neuronal correlates in response to this task seem appropriate: Almost every study using the n-back task reliably reports increasing activation changes in various but consistent cortical areas with increasing processing load, as summarized in a recent meta-analysis (Owen, McMillan, Laird, & Bullmore, 2005). Areas most commonly showing this load-dependent activation change are the bilateral dorsolateral prefrontal cortex (DLPFC) (Brodmann Area (BA) 9, 46), bilateral inferior frontal and mid-ventrolateral prefrontal regions (BA 45, 47), bilateral rostral prefrontal cortex or frontal pole (BA10), bilateral premotor regions (BA 6,8), dorsal cingulate/medial premotor cortex, including supplementary motor area (BA 32,6), bilateral and medial posterior parietal areas, including precuneus and inferior parietal lobules (BA 7, 40), and also, the medial cerebellum (Owen et al., 2005). Although these activation foci seem to be consistent irrespective of the material being processed, some regions seem to be selectively activated in accordance to the stimulus material (Nystrom et al., 2000; Owen et al., 2005).

Another line of research using the n-back task was taken by me (Jaeggi et al., 2003) in using the n-back task as *dual-task version*, but nevertheless yielding similar activation foci as described above. In my opinion, the n-back task is an even better paradigm as dual-task than as single task to challenge and even exceed the above-mentioned capacity-limits. Working mainly with students as participants it is probable that a 3-back single task proves to be an easy task in most of the cases, therefore, as dual-task condition, also students can be challenged.

Coming to task analysis, in the literature, the n-back task is generally seen as an adequate task to measure WM processes, since storage *and* manipulation, i.e., updating processes are required (e.g. Smith & Jonides, 1997) and therefore, fulfilling the requirements of a WM task (Kyllonen & Christal, 1990; Oberauer, 2005). WM in that context is mostly referred to in terms of the multicomponent model sensu Baddeley (1986) as described in Chapter 2.2. Jonides et al. (1997) suggest the following processes coming into play when performing an n-back task: Encoding (interpretation of each stimulus), storage (to retain any stimulus relevant to a future decision), rehearsal (to keep the contents of the storage active), matching (to compare every stimulus against the appropriate previous one, i.e., the one n back in the

sequence), temporal ordering (to keep the ordinal position of each stimulus tagged, that the present one can be matched), inhibition (to dampen the trace of the oldest stimulus, replacing it with the newest) and response selection, i.e., the actual execution of the response. Thus, the task seems to involve both, static storage processes (mainly in the 0- and 1-back conditions) and dynamic executive manipulation/updating processes at higher levels of load (from 2-back and above), and appears therefore to be in line with the above described modal view of WM. Applied as dual-task version, executive processing will be required at every level of load, since the allocation of attentional resources to independent stimuli is commonly referred to as core property of the CES (e.g. Baddeley, 1996).

Still, there are some drawbacks with the n-back task: Despite yielding reliable behavioural data and imaging results, the major point is that no systematic task analysis has been done apart from Jonides et al. (1997), as has been criticized recently by Meegan, Purc-Stephenson, Honsberger, and Topan (2004). But more important on theoretical level is the difficulty to separate the storage capacity from the executive component in this task, both components which are described as separate modules in Baddeley's model (Baddeley, 1986). Despite some unsolved problems, I decided to use the n-back task, trying to shed some light on the question where the capacity limits in human information processing might lie. To tackle these issues, several experiments were conducted and are now discussed in the following sections.

3.2.1. The psychometric properties of the n-back task

3.2.1.1. Reliability of the n-back task

To estimate the 'usefulness' of the n-back task in order to clarify open questions regarding capacity limitations, its psychometric properties have to be explored and described in the first place. It is astonishing that despite the huge body of research concerning the n-back task, hardly anything is reported in regard to reliability or validity of this paradigm. To my knowledge, only one study reports reliability measures (test-retest reliability) of a visuospatial version of the n-back task at four levels of load (0-3-back) yielding reliability measures between $r = 0.49$ (1-back) and $r = 0.73$ (3-back) for accuracy, and even better values for reaction times (between $r = 0.69$ (2-back) and $r = 0.86$ (0-back) (Hockey & Geffen, 2004). To get further insights into this topic, I compared reliability and stability measures (split-half and test-retest reliabilities; accuracies and RTs) in five experiments[2]. In all of these experiments I used the same n-back task with the visuospatial and auditory stimuli as shown in Figure 2. Per level of load, always 40 trials were administered; one trial consisting of the stimulus (500ms) and an interstimulus interval (ISI) of 2500 ms as (see Figure 2). The proportion of targets and non-targets was always 1:3 and the position of the targets was varied randomly. The detailed procedures are further described later in the corresponding methods sections.

The *split-half reliabilities* were assessed in Experiment 1 and 4, comparing the first half (20 trials) with the second half (20 trials) of each level of the n-back task (1 to 3-back); each half consisted of the same amount of targets (i.e., 6); however, the positions of the targets varied. The difference between the two experiments was that in Experiment 1, the 40 trials of each task level were presented as one whole run. In Experiment 4, an easier baseline task (0-back) alternated randomly with another task level every 20 trials, each level of load consisting of two separate runs comprising 20 trials. The reliability measures were assessed as Pearson's Product Moment Correlation between trials 1-20 and 21-40, and corrected with the Spearman-Brown-Prophecy formula to adjust for the underestimation of reliability compared to the reliability of the whole test (Bortz & Döring, 2002; Meier, 1999).

---

[2] These are Experiments 1, 2, 3, 6, and 7, which will be described and discussed in detail later in the corresponding chapters.

For *test-retest reliability* estimates, the results of three experiments are reported: In Experiment 2, participants performed the n-back task twice on two consecutive days, the first in our lab, and the second in the MR-scanner, with a slightly different response device (button-box vs. pneumatic squeeze balls), which should have no impact on the accuracies, but maybe on the RT-measures. As in Experiment 4, each n-back task consisted of two runs of 20 trials, alternating with the baseline task (0-back). To exclude any position effects, the position of each n-back task was determined randomly and was therefore not the same on the two occasions. In Experiments 6 and 7, participants performed the n-back task with an interval of either 10 or 20 days[3]. The test-retest reliability was assessed as Pearson's Product Moment Correlation between the two sessions.

Table 1. Split-half reliabilities for reaction time (RT) and accuracies (P_r, i.e., proportion of hits minus false alarms) in two n-back experiments (Experiment 1 and Experiment 4) shown for the various levels of load.

	Split-half reliability (RT)		Split-half reliability (P_r)	
	Experiment 1 (*N*=132)	*Experiment 4* (*N*=71)	*Experiment 1* (*N*=132)	*Experiment 4* (*N*=71)
N-back Task	r	r	r	r
Single task				
1-back visuospatial	0.94***	0.92***	0.94***	-0.08*ns*
2-back visuospatial	0.88***	0.71***	0.81***	0.21*ns*
3-back visuospatial	0.66***	0.70***	0.51***	0.55***
1-back auditory	0.90***	0.92***	-0.14*ns*	0.14*ns*
2-back auditory	0.67***	0.66***	0.58***	0.46*
3-back auditory	0.54***	0.66***	0.37**	0.38*
Dual task				
1-back visuospatial	0.80***	0.63***	0.65***	0.54**
2-back visuospatial	0.67***	0.57***	0.71***	0.69***
3-back visuospatial	0.42**	0.42*	0.62***	0.20*ns*
1-back auditory	0.88***	0.69***	0.62***	0.42*
2-back auditory	0.84***	0.64***	0.72***	0.51**
3-back auditory	0.60***	0.38*ns*	0.48***	0.54**

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ns: Not significant.
The correlation coefficients are corrected with the Spearman-Brown formula as described in the text.

---

[3] 10 days in experiment 6, and 20 days in Experiment 7; in Experiment 6 and 7, participants were systematically trained on the n-back task. For the reliability analysis, the data of the control groups were used, therefore, the pre and post measures without further intervention in-between could be compared.

Table 2. Test-Retest reliabilities for Reaction Times (RTs) and accuracies (P_r, i.e., proportion of hits minus false alarms) in three n-back experiments (Experiment 2, 6 and 7) shown for the various levels of load.

	Test-Retest reliability (RTs)		Test-retest reliability (P_r)	
	Experiment 2 (N=21)	*Experiment 6/7* (N=16)	*Experiment 2* (N=21)	*Experiment 6/7* (N=16)
N-back Task	*r*	*r*	*r*	*r*
Single task				
0-back visuospatial	0.65**	0.77***	-0.03*ns*	-0.12*ns*
1-back visuospatial	0.54*	0.84***	0.28*ns*	-0.12*ns*
2-back visuospatial	0.62**	0.83***	0.38*ns*	0.78***
3-back visuospatial	0.52*	0.38*ns*	0.50*	0.68**
0-back auditory	0.68**	0.81***	0.07*ns*	-0.13*ns*
1-back auditory	0.65**	0.76***	-0.14*ns*	0.32*ns*
2-back auditory	0.70**	0.58*	0.60**	0.11*ns*
3-back auditory	0.76**	0.68**	0.34*ns*	0.64**
Dual task				
0-back visuospatial	0.54*	0.91***	0.52*	-0.11*ns*
1-back visuospatial	0.58**	0.84***	0.57**	0.68**
2-back visuospatial	0.80**	0.57*	0.81**	0.46*ns*
3-back visuospatial	0.14*ns*	0.57*	0.71**	0.34*ns*
0-back auditory	0.50*	0.84***	0.53*	-0.12*ns*
1-back auditory	0.38*ns*	0.76***	0.35*ns*	0.04*ns*
2-back auditory	0.66**	0.28*ns*	0.86**	0.00*ns*
3-back auditory	0.21*ns*	0.41*ns*	0.60**	0.66**

Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ns: Not significant.

Tables 1 and 2 show that in general, reliability measures for the n-back task range between average and very high values. They are higher in respect of RTs, regardless of whether split-half reliability or test-retest reliability is estimated, which is consistent with reports of Hockey and Geffen (2004). The highest reliabilities for RTs are observed in the visuospatial 1-back single task. Also in RTs, the single tasks yield more consistent results across experiments and are generally higher. In the 3-back dual-task condition (auditory component), which is the most difficult condition, reliability was moderate and can be explained with the few observations of hits in this condition (since only RTs of hits are entered in the analysis while the participants' difficulty is expressed mostly with omissions).

Conversely, in accuracy, the dual tasks yield more consistent results across experiments. In the dual tasks, the largest reliability measures were observed in the 2-back tasks, which is consistent with Roznowski (1993) stating that psychometric properties such as reliability and stability improve as the task becomes more complex.

The low reliabilities in most of the easier 0- and 1-back tasks are explained with ceiling effects.

Unsurprisingly, reliability measures are most consistent in Experiment 1 (split-half reliability), where the number of participants was the largest, but still, regarding the different experimental conditions over the four experiments, it can be concluded, that despite reservations concerning reliability measures of most executive tasks (Burgess, 1997; Rabbitt, 1997), the n-back task, especially at high levels of load, can be regarded as reliable measure and can therefore be used well founded for further experiments.

The next question concerns the validity of the task and will be tackled with Experiment 1.

3.2.1.2. Validity of the n-back task

3.2.1.2.1. Associating the n-back task with other WM measures: Experiment 1

As stated before, the n-back task is usually associated and discussed in the context of WM by most of the researchers using the paradigm, and it is taken for granted, that with the n-back task, WMC is assessed. Indeed, there is some evidence for this assumption, since reliable activation in brain regions commonly associated with WM processes is observed across different groups of participants, and across different experiments and different types of material (Owen et al., 2005). Also, the processes probably taking place while executing an n-back task (storage and manipulation; Jonides et al., 1997) are commonly taken as a feature of a WM task (Oberauer, 2005). However, are these neural correlates and hypothetical task contents enough in order to classify the n-back task as WM task? To my knowledge, no systematic task analysis has been done so far, and hardly any experiment has correlated the n-back task with other WM tasks. A few studies report correlations between the n-back tasks and tasks measuring fluid intelligence (*g*) (Gevins & Smith, 2000; Gray, Chabris, & Braver, 2003; Hockey & Geffen, 2004), which will be discussed in detail in Chapter 5. The only study to my knowledge comparing the n-back task with a WM measure was published by Roberts and Gibson (2002). They compared some version of the n-back task with other WM measures, all yielding very low to moderate correlations: An $r = -0.15$ was observed between a verbal version of the n-back task and the RST (Daneman & Carpenter, 1980; see below), and $r = 0.20$, i.e., $r = 0.43$ with the forward and the backwards digit span, respectively. Yet, Roberts and Gibson (2002) used a combined measure of the n-back task (2 to 5-back, depending on the participants' performance) and therefore, it is not clear, whether the different levels of load would relate differentially to other measures of WM. In my opinion, combining the 2 to 5-back levels into a single measure of n-back is not helpful to clarify the relationship between the n-back task and other WM measures, especially if not every participant does the same levels of load, as in the study by Roberts and Gibson (2002). It would be preferable to use the n-back task with various type of material and at various levels of load comparing it with other measures of WM, and also to test whether material- or modality-specific processes play a

differential part, or whether more executive, modality-free processes in the sense of Engle, Tuholski, Laughlin, and Conway (1999) are tapped.

Following this line of arguments, in Experiment 1, different levels of load of the n-back task were compared with a typical measure of WMC (i.e., the RST). According to Daneman and Merikle (1996), Just and Carpenter (1992), as well as Whitney, Arnett, Driver, and Budd (2001), the RST is one of the most commonly used assessments of WMC and especially of executive functions, making it therefore the ideal task to compare with the n-back task, which aims to measure comparable processes. The RST consists of 100 relatively simple and unrelated sentences (with 6 additional training sentences), which participants have to read aloud and indicate with 'yes' or 'no' for each sentence, whether it makes (semantically) sense or not. Sentences are removed, as soon as the participants have made their yes/no decision. Additionally, participants have to retain the last word of each sentence and recall these words in the correct order after two, three, four, five, or six sentences. In the RST, there are many processes involved, i.e., the 'usual demands of sentence comprehension, from the lower level processes that encode the visual patterns of individual words and access their meanings, to the higher level processes that compute the semantic, syntactic, and referential relations among the successive words (…)' (Daneman & Merikle, 1996; p. 423). Further, additional and simultaneous resources are needed in order to recall the final words, such as rehearsal and time sharing processes. Therefore, the RST can be regarded as 'dual task' and certainly executive to some extent since attention has to be divided between various processes, and consequently, I expected substantial correlations between the RST and the dual-task version of the n-back task, especially at 2- and 3-back levels (accuracy). Moreover, I expected the auditory versions of the dual-tasks to be higher correlated with the RST than the visuospatial versions, since in the auditory version, more verbal processes are present, which might be processes shared also in the RST.

3.2.1.2.2. Method

Participants. 132 participants took part in the study without any specified selection criteria apart from being natively German speakers. 65 were women and 67 were men; and the mean age was 27.93 (*SD* = 10.28; *Range* = 14-67).

Material and Procedure. The n-back task was used at 1 to 3-back levels with visuospatial and verbal material (see Figure 2), administered both as single and dual tasks. The visual-nonverbal stimuli consisted of blue squares, appearing in eight different loci on the computer screen. The auditory material comprised eight German consonants (c, g, h, k, p, q, t, w) spoken in a female voice and selected on the basis of their distinctiveness. Stimuli were presented for 500 ms with an interstimulus interval (ISI) of 2500 ms consisting of a white fixation cross on a black screen. In the dual-task version, the visual-nonverbal and the auditory-verbal stimuli were presented simultaneously. Each task-load condition consisted of 40 trials with 12 of them being targets (i.e., 24 in the dual-task condition; 12 in each modality), which were at randomly assigned positions. Participants responded with their left index finger on the leftmost button on a button-box in regard to visuospatial material, and with their right index finger on the rightmost button in regard to the auditory material.

Before starting with the n-back task, a short practice session was administered in order to familiarize the participants with the tasks and to make sure that each task was properly understood (normally, 10 trials per level of load were sufficient; but if necessary and requested by the participants, some further practice trials were added). After this practice session, participants first completed the single tasks (half of the subjects started with the visual-nonverbal task, and half of them started with the auditory-verbal task) and followed by the dual task, always starting with the 1-back task, followed by the 2-back and the 3-back task.

RST. The material for the RST was provided by courtesy of Meredyth Daneman and translated into German[4]. Each of the 100 sentences contains 6 to 15 words (*M*: 10.05; *SD*: 1.98) with a mean word length of 6.25 (*SD*: 0.81). Examples: 'The young woman sang the empty blue bottle' (makes no sense, therefore 'no'); 'He pulled up a chair and joined the card game' (makes sense, therefore 'yes'). The final words to be recalled in this example would be 'bottle' and 'game'. Half of the sen-

---

[4] The translation was undertaken by myself and undergraduate students during the experimental-practical training course in 2003/04.

tences make sense and half of them do not make sense, but all of them are syntactically correct. The complete material (in German) can be found in the appendix. The measure of performance (i.e., the reading span) is defined as the maximum number of sentences that the participants are able to process, while maintaining perfect recall of the final words in more than 50%. This measure serves as index of WMC (Daneman & Merikle, 1996; Just & Carpenter, 1992) and also corresponds to the criterion to stop with the task, also in order not to yield too much frustration in the participants. Participants' performance was monitored in two ways: First, the above mentioned traditional span score was assessed, and second, the number of correctly recalled words across all performed sentence sets was collected, in order to obtain a more continuous score, and therefore, according to Saito and Miyake (2004) being psychometrically more appropriate.

Further, a task for crystalline, verbal intelligence was administered, which will be described later in Experiment 5. Also, a prospective memory task was embedded in a visuospatial choice RT-task at the end of the experiment session, which will be reported somewhere else (Jaeggi, Meier, & Buschkuehl, in prep.).

Analysis. For the n-back task, 3-way repeated measures ANOVAs with load (1 to 3-back) x task (single vs. dual) x modality (visuospatial vs. verbal) as within-subject factors were analyzed for accuracy as assessed with the discrimination index P_r (hits minus false alarms) according to Snodgrass and Corwin (1988)[5], and for RTs as dependent variables. Post-hoc tests were applied in case of significant interactions, and non-parametric pairwise comparisons were calculated in case of non-normal distributions.

Further, the means and standard-deviations of both, the n-back task (1-3-back; visuospatial and auditory-verbal, single and dual tasks) and the RST (span-measure and the number of correct responses), were used as follows: Pearson's Product Moment Correlations were calculated between each level of the n-back task and the RST in order to estimate the inter-correlations between tasks, as well as the relationship between the two WM measures.

---

[5] This measure varies between 0 and 1, whereas 1 reflects perfect performance, and 0 equals performance at chance level.

3.2.1.2.3. Results

n-back task. Means and standard deviations concerning the performance measures (P_r and RTs) are shown in Table 3. The 3-way repeated measures analysis of variance (ANOVA) yielded a main effect of load (1- to 3-back; P_r: $F(2,131) = 1430.42$; $p < 0.001$; RTs: $F(1.58,131) = 207.33$; $p < 0.001$), a main effect of task (single vs. dual; P_r: $F(1,131) = 462.80$; $p < 0.001$; RTs: $F(1,131) = 627.89$; $p < 0.001$), and a main effect of modality (visuospatial vs. auditory; P_r: $F(1,131) = 29.59$; $p < 0.001$; RTs: $F(1,131) = 193.53$; $p < 0.001$), as well as significant interactions for load x task (P_r: $F(1.80,131) = 69.52$; $p < 0.001$; RTs: $F(1.60,131) = 19.88$; $p < 0.001$), load x modality (P_r: $F(1.90,131) = 30.98$; $p < 0.001$; RTs: $F(1.51,131) = 40.81$; $p < 0.001$), and for task x modality (P_r: $F(1,131) = 6.18$; $p < 0.05$; RTs: $F(1,131) = 8.70$; $p < 0.01$). The 3-way interaction load x task x modality was also significant (P_r: $F(1.65,131) = 6.77$; $p < 0.01$; RTs: $F(1.46,131) = 5.69$; $p < 0.01$).

Post-hoc tests revealed that the difference between the single and the dual tasks increased with increasing load and was most expressed at the 3-back level (*1-back*: $Z = -7.58$; $p < 0.001$; *2-back*: $Z = -9.40$; $p < 0.001$; *3-back*: $Z = -9.79$; $p < 0.001$), where the performance was lowest in the dual-task. Similarly, the lower performance in the auditory than in the visuospatial tasks was more pronounced at higher levels of load with a significant difference at the 3-back level ($Z = -6.44$; $p < 0.001$). However, modality had a larger effect in the single tasks, i.e., the above-mentioned lower performance in the auditory tasks was more pronounced in the single tasks than in the dual tasks (*single tasks*: $Z = 6.27$; $p < 0.001$; *dual tasks*: $Z = -2.33$; $p < 0.001$). The 3-way interaction can be similarly interpreted in the differential role that modality plays at higher levels of load and in the single tasks compared to the dual tasks: Participants performed better in the visuospatial tasks, especially at higher levels of load, as being most expressed in the 3-back single task (*1-back single*: $Z = -3.25$; $p = 0.001$; *3-back single*: $Z = -6.88$; $p < 0.001$; *3-back dual*: $Z = -3.06$; $p < 0.01$). Almost the same applied to the RTs: Participants reacted faster in the single-task conditions than in the dual tasks, but this time, the difference was smallest in the 3-back condition (*1-back*: $t(131) = -22.73$; $p < 0.001$; *2-back*: $t(130) = 25.44$; $p < 0.001$; *3-back*: $t(126) = -16.18$; $p < 0.001$). The visuospatial tasks yielded faster response times than the auditory tasks, with the largest difference found in the 3-back condition (*1-back*: $t(131) = -11.21$; $p < 0.001$; *2-back*: $t(130) = -9.85$; $p < 0.001$;

3-back: $t(126) = -12.48$; $p < 0.001$). This difference between modalities was more pronounced in the dual tasks (*single tasks*: $t(131) = 14.78$; $p < 0.001$; *dual tasks*: $t(126) = 9.75$; $p < 0.001$). The 3-way interaction indicated that the faster responses in the visuospatial tasks were more expressed at higher levels of load in the single tasks, with the largest difference observed at the 3-back level, but conversely, at lower load levels in the dual tasks (single tasks: *1-back*: $t(131) = -7.99$; $p < 0.001$; *2-back*: $t(131) = -8.99$; $p < 0.001$; *3-back*: $t(131) = -12.20$; $p < 0.001$; dual tasks: *1-back*: $t(131) = -9.49$; $p < 0.001$; *2-back*: $t(130) = -7.38$; $p < 0.001$; *3-back*: $t(126) = -7.83$; $p < 0.001$).

RST. Performance measures, i.e., the span score and the total number of correctly recalled target words are also reported in Table 3. The correlation between the classic span measure and the number of correctly recalled items was $r = 0.91$ ($p < 0.001$), indicating therefore that depending on the research-interests and preferences, either measure can be taken as equivalent measure of WMC.

Table 3. Means (*M*) and standard deviations (*SD*) for the various performance measures (*N* = 132).

		P_r		RTs (ms)	
N-back Task		*M*	*SD*	*M*	*SD*
Single tasks	1-back visuospatial	0.99	0.07	514	175
	2-back visuospatial	0.89	0.14	550	181
	3-back visuospatial	0.64	0.19	672	253
	1-back auditory	0.98	0.05	596	138
	2-back auditory	0.88	0.13	697	192
	3-back auditory	0.47	0.19	1049	371
Dual tasks	1-back visuospatial	0.93	0.13	934	262
	2-back visuospatial	0.71	0.22	1159	350
	3-back visuospatial	0.35	0.20	1225	448
	1-back auditory	0.92	0.12	1174	349
	2-back auditory	0.69	0.23	1429	435
	3-back auditory	0.27	0.19	1610	531
Reading Span Task (RST)					
	number of correctly remembered words	36.42	15.64		
	span	2.75	0.83		

Note. For the n-back task, accuracies are described as P_r (i.e., proportion hits minus false alarms). Reaction times (RTs) are reported as average median values from hits only. For the RST, no RTs were recorded, and accuracies are reported as means of correctly remembered words as well as the span measure as described in the text.

Correlations: The intercorrelations between the various levels of the n-back are plotted on Table 4. It is evident that the different n-back conditions are moderately related and more consistently expressed in terms of RTs. But also for accuracies, some correlations are substantial, usually more pronounced at lower levels of load. The low correlations between the accuracy and the speed-measures show that there is no speed-accuracy tradeoff in the various levels of the n-back task, apart from the single auditory 2-back score (r = -0.41; $p < 0.001$), where the negative correlation indicates that participants with slower response latencies also made more errors.

Table 4. Intercorrelations between the various n-back tasks for reaction time (RT) and accuracy, i.e., hits minus false alarms (P_r).

		RT: Single Tasks						RT: Dual Tasks						P_r: Single Tasks						P_r: Dual Tasks					
		visuospatial			auditory			visuospatial			auditory			visuospatial			auditory			visuospatial			auditory		
RT: Single Tasks		1back	2back	3back	1back	2back	3back	1back	2back	3back	1back	2back	3back	1back	2back	3back	1back	2back	3back	1back	2back	3back	1back	2back	3back
visuospatial	1back	1																							
	2back	0.66**	1																						
	3back	0.52**	0.55**	1																					
auditory	1back	0.74**	0.55**	0.41**	1																				
	2back	0.50**	0.49**	0.43**	0.59**	1																			
	3back	0.25**	0.32**	0.40**	0.30**	0.38**	1																		
RT: Dual Tasks																									
visuospatial	1back	0.30**	0.39**	0.28**	0.37**	0.44**	0.33**	1																	
	2back	0.19*	0.41**	0.38**	0.29**	0.35**	0.31**	0.64**	1																
	3back	0.12	0.29**	0.32**	0.24**	0.26**	0.18*	0.52**	0.52**	1															
auditory	1back	0.26**	0.21*	0.24**	0.42**	0.44**	0.27**	0.58**	0.43**	0.39**	1														
	2back	0.14	0.21*	0.25**	0.28**	0.32**	0.37**	0.42**	0.47**	0.38**	0.77**	1													
	3back	0.05	0.19*	0.32**	0.18*	0.28**	0.30**	0.33**	0.38**	0.40**	0.66**	0.66**	1												
P_r: Single Tasks																									
visuospatial	1back	0.01	-0.08	0.03	0.01	0.11	0.01	0.06	0.06	0.08	0.16	0.13	0.18*	1											
	2back	0.13	0.07	0.05	0.08	0.00	0.08	0.02	-0.03	0.07	0.18*	0.17	0.25**	0.37**	1										
	3back	-0.16	-0.19*	-0.12	-0.17*	-0.10	-0.11	-0.13	-0.07	-0.18*	-0.05	-0.04	0.08	0.14	0.35**	1									
auditory	1back	-0.04	-0.02	0.14	-0.02	0.07	0.00	-0.02	-0.01	0.04	0.17	0.19*	0.24**	0.63**	0.40**	0.21*	1								
	2back	-0.02	0.00	-0.03	-0.14	-0.41**	-0.20*	-0.13	-0.12	-0.08	-0.12	-0.07	0.07	0.08	0.15	0.18*	0.16	1							
	3back	0.02	0.03	0.05	-0.08	-0.19*	0.01	-0.18*	-0.11	0.00	0.01	0.01	0.09	0.10	0.21*	0.18*	0.05	0.31**	1						
P_r: Dual Tasks																									
visuospatial	1back	-0.01	0.01	0.10	-0.10	-0.04	-0.06	-0.15	0.09	0.10	-0.06	-0.03	0.13	0.26**	0.20*	0.17	0.34**	0.20*	0.20*	1					
	2back	0.08	0.03	0.12	0.07	0.09	0.15	-0.10	-0.01	0.08	0.12	0.16	0.27**	0.28**	0.48**	0.39**	0.44**	0.21*	0.24**	0.47**	1				
	3back	0.05	0.06	0.12	0.18*	0.07	0.05	-0.15	0.01	0.11	0.13	0.16	0.20*	0.18*	0.27**	0.28**	0.31**	0.14	0.19*	0.27**	0.46**	1			
auditory	1back	-0.07	-0.05	-0.01	-0.16	-0.06	-0.10	-0.10	0.02	0.04	-0.12	-0.12	-0.08	0.26**	0.08	0.07	0.29**	0.22**	0.18*	0.65**	0.33**	0.13	1		
	2back	-0.05	0.05	0.08	-0.09	-0.11	0.08	-0.09	0.01	-0.02	0.03	-0.01	0.15	0.12	0.21*	0.33**	0.20*	0.40**	0.41**	0.18*	0.34**	0.24**	0.20	1	
	3back	0.00	0.00	0.01	-0.14	-0.19*	-0.14	-0.06	-0.10	-0.08	-0.07	-0.08	0.00	0.05	0.08	0.29**	0.03	0.31**	0.37**	0.27**	0.16	0.05	0.25**	0.36**	1

Note: N = 132; RT consist of the average median values from hits only (ms). * $p < 0.05$; ** $p < 0.01$.

The correlations between the various levels of the n-back task and the RST were very low, if existent at all (see Tables 5 and 6): The only significant correlation concerning the accuracy measures (P_r) was observed between the RST (number of correct responses) and the auditory 2-back dual task ($r = 0.17$; $p = 0.05$). Correlating the RTs at different levels of load with the RST, the 3-back single tasks (auditory and visuospatial) were significantly related to the RST, but with $r = 0.22$ ($p < 0.05$) for the auditory task and $r = 0.18$ and $r = 0.20$ ($p < 0.05$) for the visuospatial tasks, the correlations were also marginal. Similarly, as with the accuracy measures, a significant correlation was observed between the auditory dual 2-back task and the RST ($r = -0.17$; $p < 0.05$). Yet, as with the other correlations, the value is only small and only reaching significance if not the span, but the number of remembered words was taken as RST measure. If composite measures were calculated and taken as measure to correlate with the RST as done by Roberts & Gibson (2002), no substantial correlations between either RST measure and the n-back tasks were observed, which is in accordance with the findings of Roberts & Gibson (2002). The only significant correlation was found between the 'more executive' auditory dual-tasks (2 and 3-back; $r = 0.20$; $p < 0.05$), or if all auditory dual tasks (1 to 3-back) were combined ($r = 0.23$; $p < 0.01$), and therefore, not substantially improving the values of the individual load measures.

Table 5. Pearson's Product Moment correlations between the two measures of the RST and the *accuracy measures* (P_r) in the various levels of the n-back task ($N = 132$).

	RST (remembered words)	RST (span measure)
N-back Task		
Single Tasks		
1-back auditory	0.02	0.04
2-back auditory	-0.02	0.01
3-back auditory	0.22*	0.22*
1-back visuospatial	0.05	0.06
2-back visuospatial	0.10	0.08
3-back visuospatial	0.18*	0.20*
Dual Tasks		
1-back auditory	-0.08	-0.04
2-back auditory	-0.17*	-0.14
3-back auditory	-0.08	-0.07
1-back visuospatial	0.12	0.12
2-back visuospatial	0.03	0.02
3-back visuospatial	0.00	0.04

*Note. * $p < 0.05$ (2-tailed).

Table 6. Pearson's Product Moment correlations between the two measures of the RST and the *RTs* in the various levels of the n-back task ($N = 132$).

	RST (remembered words)	RST (span measure)
N-back Task		
Single Tasks		
1-back auditory	-0.05	-0.01
2-back auditory	-0.05	-0.07
3-back auditory	0.08	0.07
1-back visuospatial	-0.05	0.00
2-back visuospatial	-0.02	-0.02
3-back visuospatial	-0.04	-0.02
Dual Tasks		
1-back auditory	0.12	0.12
2-back auditory	0.17*	0.15
3-back auditory	0.16	0.10
1-back visuospatial	0.14	0.12
2-back visuospatial	-0.03	-0.04
3-back visuospatial	-0.02	0.04

*Note. * $p < 0.05$ (2-tailed).

3.2.1.2.4. Discussion

Regarding the n-back task, my results provide evidence, that the load manipulation was successful in both modalities, but more expressed in the auditory modality as indicated with the longer response latencies and more errors, especially at high levels of load. Even more demands are placed on the processing system, if the task is conducted as dual task as indicated with the interaction of load x task (single vs. dual), again more pronounced in the auditory modality as expressed in the significant 3-way interaction (load x task x modality).

The intercorrelations between the various n-back conditions show that despite the various load and modality-conditions the task nevertheless seems to be measuring related processing mechanisms. No correlations between speed and accuracy scores were observed, apart from the auditory single 2-back task, indicating that there was no speed-accuracy trade-off, which is consistent with the findings of Hockey and Geffen (2004).

Correlation analyses between the two measures do not provide much evidence that the n-back task and the RST are related to the same construct, such as for example, WM. Therefore, it can be argued that either the RST or the n-back task does not measure WM. However, the RST is a long and well established measure and uniquely referenced in the context of WM, despite some critical voices (e.g. Whitney et al., 2001), stating that there are far more processes involved in a RST than 'mere' WM, such as susceptibility to interference. Nevertheless, it seems that the n-back task has some unexpected source of variance not easily explained with the traditional views of WM or the CES. The most probable explanation would be that WM and/or the central executive are not unitary and that the RST and the n-back task are not loading on the same WM or executive sub-component. Evidence that these tasks might share some communalities, despite their low inter-relationship concerning behavioural data, comes from neuroimaging studies: In the brain, overlapping areas, mainly prefrontal, are activated when either the RST (Osaka et al., 2004), or the n-back task (Jaeggi et al., 2003; Wager & Smith, 2003) is performed, so it is likely, that some processes are shared in the two tasks. Maybe both tasks are involved in processes to be subsumed as 'executive', such as attention allocation, providing further evidence in favour of a multimodal view of WM and executive functions (Miyake et al., 2000; Salthouse, Atkinson, & Berish, 2003; Stuss et al., 2002).

Alternatively, the low spatial resolution in fMRI and PET might prevent findings of smaller regions within the prefrontal areas, which are differentially activated in either task and therefore reflecting different processes; an explanation discussed for example in Cabeza and Nyberg (2000; 2003) as 'within regions-approach'.

Still, to summarize, the question concerning the validity of the n-back task cannot be answered for certain with Experiment 1 and has to be further explored, also in the forthcoming experiments. Coming back to the RST, which is predictive for a wide range of 'higher order' abilities, such as reading comprehension, it will be interesting to see, whether with the n-back task other tasks can be predicted, like for example fluid intelligence (g); a relation which has be shown already by other researchers (e.g. Gevins & Smith, 2000; Gray et al., 2003), and which will be discussed in Chapter 5. However, before investigating the relationship of the n-back task with intelligence, I will discuss the neurophysiological basis of capacity limitations.

4. Capacity limitations and neurophysiology – The view of neuroimaging

4.1. General issues

For a few years now, the interest in capacity limitations has also been growing in the domain of neurophysiology and functional neuroimaging. Still, only a few studies have addressed the question regarding the neurophysiological consequences if capacity limitations in information processing are reached or exceeded (e.g. Callicott et al., 1999; Goldberg et al., 1998; Goldman-Rakic, 1996; Handy, 2000; Jaeggi et al., 2003; Klingberg, 2000; Ramsey, Jansma, Jager, Van Raalten, & Kahn, 2004; Rypma, Berger, & D'Esposito, 2002). Some of them use single tasks at various levels of load to observe brain activity at the edge of capacity (e.g. Callicott et al., 1999), while others are concerned with the exploration of the neural correlates of the central executive system using dual-task paradigms (e.g. D'Esposito et al., 1995). In the literature, there is some inconsistency and controversy as to the neural correlates of dual-task processing. While some researchers observed a dual-task specific area, first reported by D'Esposito et al. (1995), and more recently by Kondo, Osaka, and Osaka (2004), and Collette et al. (2005), other researchers found no qualitative difference between single- and dual-task processing (Jaeggi et al., 2003), however, more activation was observed compared to single-task processing (Adcock, Constable, Gore, & Goldman-Rakic, 2000; Bunge, Klingberg, Jacobsen, & Gabrieli, 2000; Klingberg & Roland, 1997), with some research groups even reporting activation decreases in dual-task processing (Fletcher et al., 1995; Goldberg et al., 1998; Just et al., 2001).

Trying to integrate these intriguing results, Klingberg (1998; 2000) proposed three possible neurophysiological mechanisms being linked to dual-task performance and related performance decreases: a) a specific dual-task area, b) overlapping fields in the cortex, and c) cross-modal inhibition processes.

a) The first assumption, that dual-tasking can be *qualitatively* differentiated from single task processing can be summarized as follows: More complex and different operations, such as task-coordination are required when simultaneously performing two tasks, resulting in the recruitment of additional cortical areas in the sense of a 'coordination-module'. Performance decreases could be explained with sub-optimal functioning of this coordination-module in some circumstances. Evidence

for such a sub-functioning comes from studies with patients with Alzheimer's disease (AD; e.g. Logie, Cocchini, Della Sala, & Baddeley, 2004), showing that AD-patients are disproportionally impaired in performing dual-tasks, compared to difficulty-matched single-tasks. In functional neuroimaging, the above mentioned and widely cited study by D'Esposito and colleagues (1995) reported a 'surplus' activation in the DLPFC, which was only observed in the dual-task conditions and not during the performance of the two single tasks alone, a finding recently supported by other research teams (Collette et al., 2005; Kondo et al., 2004). This activation focus in the DLPFC during simultaneous processing was taken as evidence for a special area only recruited in dual-tasking and as possible neural correlate of the CES, since the single tasks were especially chosen not to recruit the CES and prefrontal areas per se, also having different input modalities (i.e., a simple spatial relations judgment and a semantic category judgment). However, these results could not be replicated by several other researchers (e.g. Adcock et al., 2000; Bunge et al., 2000; Jaeggi et al., 2003; Just et al., 2001; Klingberg, 1998; Klingberg & Roland, 1997), not even if exactly the same tasks were used (Adcock et al., 1999, as cited in Klingberg, 2000), however, *more* activation was observed in those areas, already activated in the single tasks. At present and despite the two recent studies (Collette et al., 2005; Kondo et al., 2004), it seems that the majority of functional neuroimaging studies failed to detect a separate 'dual-task module', and it has to be concluded that other factors than the task itself were responsible for this 'surplus' activation. In the view of Klingberg (2000), the statistical threshold (to indicate a significant signal change) is a likely candidate to explain the finding of a dual-task specific area by D'Esposito et al. (1995). On the other hand, the most recent findings by Kondo et al. (2004) and Collette et al. (2005) raise doubts about this explanation and the controversy in the literature might continue.

b) The second mechanism causing performance deterioration in dual-task situations is in line with common resource theories (e.g. Broadbent, 1957; Friedman, Campbell Polson, Dafoe, & Gaskill, 1982; Kahneman, 1973; Wickens, 1980): If two tasks compete for the same resource pool, i.e., in neurophysiological terms, for the same or overlapping parts of the cortex at the same time, the resulting pattern should be interference manifested in a performance decrease. Klingberg and Roland (1997)

suggest an *overlap-hypothesis*, also allowing predictions to the degree of interference based on the following assumptions:

'(…) if task A and B have more overlap of activation than task C and D, then there is also more interference between task A and B than between task C and D when the tasks are performed simultaneously. The reason why two tasks would interfere if they produce overlapping activations could be either that they depend on activation of the very same columns in the overlapping region, or that they activate different but inter-digitizing populations of columns, which inhibit each other by interneurons (Juliano et al., 1989; Welker et al., 1993).' (Klingberg & Roland, 1997; p. 98).

In turn, if a performance decrease is observed even though tasks do not activate overlapping fields, the overlap-hypothesis could be rejected. Moreover, predictions about the interference could be made, provided that the involved cortical areas are known. There is some evidence in the literature contributing to the overlap-hypothesis, for example by Klingberg himself (1998), Jaeggi et al (2003), but also by others (e.g. Adcock et al., 2000; Bunge et al., 2000; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999).

c) The last factor, possibly contributing to a performance decrease in dual-task situations again corresponds to Wickens' model (1984) and states, that performance decreases if there are some cross-modal *inhibition* processes causing interferences between two tasks involving stimuli from different sensory modalities. To date, there is only weak evidence for this hypothesis. Indirect support comes from the work of Just, Carpenter and collaborators (Just et al., 2001): With a sentence comprehension and a mental rotation task, they report less neuronal activation in a dual-task condition (in sensory, but also in association areas), compared to the activation during the corresponding single tasks. These results were interpreted in constraints placed on the system if faced with a dual-task condition. These constraints may be of biological basis and represent the limit of the amount of cortical tissue which can be activated at any given time, being in turn reflected behaviourally in attentional limitations, i.e., in decreased performance. The two other studies reporting decreases in the prefrontal cortex in dual-task compared to single-task conditions, i.e., Goldberg et al. (1998) combining the Wisconsin Card Sorting Task (WCST) with a verbal task, and Fletcher et al. (1995), combining an episodic memory task with a motor task, are interpreted as

possible evidence for inhibition processes by Klingberg (2000). Still, other explanations may be possible, such as inhibition processes taking place if two processes depend on a common resource pool, but not because of cross-modal interference, but because of overlapping systems as described in hypothesis b). Alternatively, a most natural hemodynamic phenomenon is discussed to take place in conditions with high processing demands, such as a saturation of the BOLD response (Friston, Mechelli, Turner, & Price, 2000).

To summarize, so far the literature concerning the neural basis of dual-task processing remains inconsistent. Most results seem to be in favour of hypothesis b), i.e., not speaking for a special module being active only in dual-tasking, and also supporting a modality-free processing system. Further, whether or not prefrontal activity is observed in the dual-task condition seems to depend on the activity present in the single-task conditions: If no (or rather, small) activation is observed in the single-task conditions, the activation in the dual-task condition is likely to be strong in prefrontal regions; if the single-tasks per se recruit a reasonable amount of prefrontal areas (such as for example in Goldberg et al., 1998), the activation might decrease in the dual-task. Adcock et al. (2000), as well as Goldman-Rakic (1996) however argue that executive processes may be mediated by interactions between anatomically and functionally distinct systems engaged in performance of component tasks, as opposed to an area or areas dedicated to a generic executive system. That is, the assumed CES can be merely and most parsimoniously interpreted as '(…) an emergent property of coactivated multiple domain-specific processors located in prefrontal cortex but interconnected both with the domain-relevant long-term storage sites in posterior regions of the cortex (sensory) and with appropriate motor pathways.' (Goldman-Rakic, 1996; p. 1450). There is no necessity to postulate a separate module, being recruited for example in dual-task situations, rather, the *interaction* between different brain regions seems to be relevant; a notion which seems very reasonable to me.

It is also far from clear which processes are taking place in single tasks if capacity limitations are reached, as investigated by several research teams (e.g. Braver et al., 1997; e.g. Callicott et al., 1999; Jansma, Ramsey, Coppola, & Kahn, 2000; Rypma et al., 2002; Rypma & D'Esposito, 1999). It has been argued, that the function of task load is directly related to activation increases in cortical regions and directly reflecting resource limitations on neuronal basis (Braver et al., 1997; Carpenter, Just,

Keller, Eddy, & Thulborn, 1999; Handy, 2000). However, there has been some evidence, that capacity limits show themselves in decreased activity in the involved areas (Callicott et al., 1999), leading to the hypothesis, that the performance curve cannot easily be applied according to activation patterns. Handy (2000) argues that three predictions of function-related covariation can be made following the assumed relationship between load and activation intensity (in single tasks): 'a) when the load on a process is manipulated, positive covariation between cortical loci suggests a sharing of processing capacity, b) negative covariation suggests a competition for capacity, and c) no covariation suggests an independence of capacities.' (Handy, 2000; p. 1066). There are different levels, where such covariations could take place, for example at molecular level (i.e., neurotransmitters), at receptive field level, or at columnar levels. Whether the proposed capacity limitations at the various levels may apply separately or interdependently is not easily answered at this stage of research, but in the view of Handy (2000), 'the capacity based approach provides the necessary theoretical structure for integrating concepts across multiple levels of empirical analysis.' (p.1068).
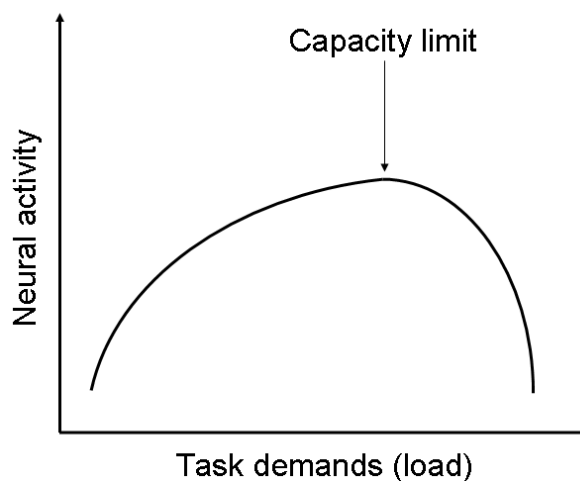
Figure 3. Hypothetical curve representing the proposed activation changes in dual-tasking formulated by D'Esposito (2001); see text for further explanation.

The remaining question not addressed yet concerns the performance at supra-capacity level. What would happen on neural and functional level, if participants perform a task beyond their capacity limitation? There have been very few studies in functional neuroimaging measuring brain activation at the edge of capacity and at supracapacity levels, so for example Goldberg et al. (1998). However, integrating the

inconsistent results from the various dual-task studies, D'Esposito (2001) formulated a hypothesis about what might happen, if capacity limitations are reached and exceeded in a dual-task situation: First, an activation increase should be observed, mainly in prefrontal areas, but if capacity limits are reached, the activation should reach a sort of a plateau before attenuating when capacity levels are exceeded, maybe because of a breakdown of neural networks. The activation pattern in prefrontal areas should therefore resemble an inverted u-shaped curve (see Figure 3). This hypothesis caught my interest and was the main reason, I conducted the study, already mentioned earlier (Jaeggi et al., 2003), and I will briefly report the main findings concerning this u-shaped curve in the following section.

4.2. Breakdown of neural networks at capacity limitations?

This experiment, consisting of my 'Lizentiatsarbeit' was the starting point for my following fMRI-experiments. Primarily, I was interested to see whether there are observable differences at neural level, if participants perform a *single* or a *dual* task at different levels of load, thus testing the hypothesis to the existence of a separate 'dual-task module'. Further, I wanted to test the above-mentioned hypothesis concerning the activation patterns in dual tasks as proposed by D'Esposito (2001), that is, whether an inverse u-shaped curve would be present in regard to the activation patterns in the dual-tasks.

Six normal, right-handed and young participants took part in the study performing the parametric n-back task at four levels of difficulty (0 to 3-back) as single and as dual task as described earlier. However, instead of the visuospatial material as in Experiment 1, 10 abstract shapes (Attneave & Arnoult, 1956) were used, and for the auditory-verbal material, 10 spoken consonants were presented through headphones.

Behavioural data as well as functional data were assessed and the results can be summarized as follows: In prefrontal cortices, there was a load-dependent activation change, i.e., the higher the load, the higher the activation. However, no activation differences between single and dual tasks were observed, and an activation increase was still observed in the dual tasks, despite the performance being near chance level. This pattern applied to several brain areas apart from the bilateral DLPFC, i.e., to the area surrounding the precentral sulcus, superior parietal areas, and to some extent, to inferior frontal regions (see Figure 4).
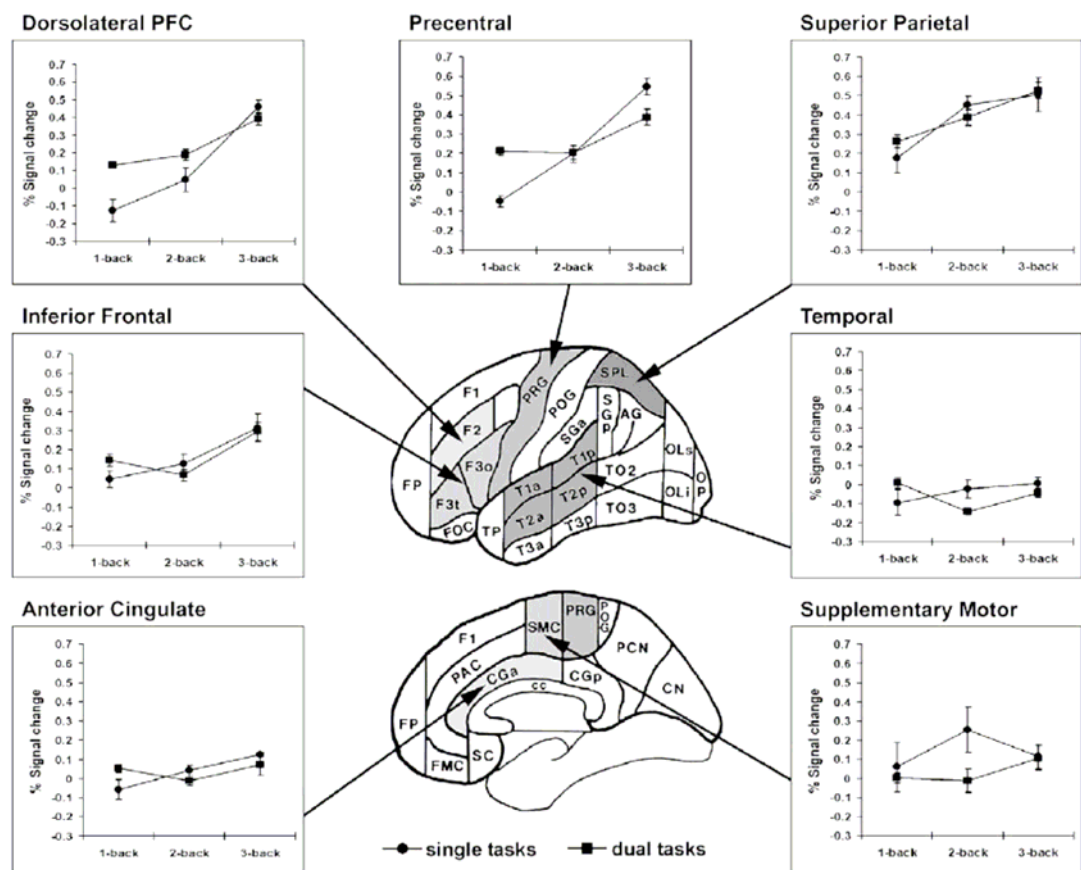
Figure 4. The schematic drawing in the centre shows the anatomical defined volumes of interest (VOI) investigated in our study, which were adapted from the parcellation scheme described and depicted in Rademacher, Galaburda, Kennedy, Filipek, and Caviness (1992). Laterally, the dorsolateral prefrontal cortex, the inferior frontal gyrus, the precentral gyrus, the superior parietal lobule, and the temporal region were defined, and medially, the supplementary motor area, and the anterior cingulate gyrus. Each area shaded in grey represents one VOI. The associated graphs depict the amount of activation in the seven given VOIs as a function of memory load (1-back to 3-back tasks, with the 0-back task as control) for both, the single and the dual tasks. The amount of activation is defined as group average of percentage BOLD (blood oxygenation level dependent) signal change. The error bars represent the standard error of the mean (From Jaeggi et al., 2003).

Thus, neither evidence for a dual-task specific module could be provided, nor an inverted u-shaped curve could be demonstrated with our data. Nevertheless, the results were very interesting and opened a wide range of possible lines of research. The primary question was now on how the obtained results were best interpreted and how the BOLD-signal might interact with the behavioural data. The capacity limitations were obviously reached as was indicated by the behavioural data; therefore the hypothesis that an activation decrease would take place at higher levels of load does not seem to apply. What might be the reason for that finding? Since variability

was high in the dual tasks, an appealing possibility would be to investigate interindividual differences in response to task difficulty and the related BOLD signal changes. There is evidence that high- and low-performing participants react differently on neuronal level if confronted with difficult material as was indicated with findings by Rypma and collaborators (Rypma et al., 2002; Rypma & D'Esposito, 1999). Following this line of argument, it might be possible that the inverted u-shaped curve as proposed by D'Esposito (2001) might only apply at certain circumstances and/or in a particular group of participants: The participants in Jaeggi et al. (2003) were all highly motivated and, as was reported by them, tried to succeed in the task, even though the task was very difficult. Therefore, it can be concluded that the attentional allocation and mental effort could be the reason for the still increasing activation in the most difficult 3-back dual task. The neural circuits adopted by prefrontal areas might be the same or at least highly overlapping in both WM or effort-related processing and therefore yielding the observable load-dependent activation patterns. The question was now, whether these increasing activation patterns would change, if low-performing and/or less motivated participants would be scanned. Therefore, an interindividual-differences approach was considered as fruitful in order to shed light on these questions and will be discussed in the following chapter.

4.3. Interindividual differences of capacity as assessed with fMRI

It seems difficult to reach to an understanding and a theory of capacity limitations without considering and understanding the basis of widely observed and marked interindividual differences and age-related changes in these limitations. Differences in processing capacity, for example in the domain of WM as assessed with the RST are predictive for a wide range of higher order cognitive tasks, such as language comprehension, problem solving, and reasoning, as demonstrated for example by Daneman and Carpenter (1980). Differences in WM performance are also related to aging processes (e.g. Meguro et al., 2000; Park et al., 2002). While there are associations between age-related neuronal changes and behavioural dysfunction (e.g. Della-Maggiore, Grady, & McIntosh, 2002), it seems likely that also individual differences in young adults might result in differential patterns in the brain. There are several questions to be asked: Specifically, are there differential activation patterns in high- and low-performing participants? Are such activation patterns directly related to the participants' performance? Do we observe differential activation patterns in high- and low-performing participants at the edge of capacity? How are older individuals' activation patterns related to those of young participants? In order to shed some light on these questions, I conducted the following three experiments with fMRI, which will be described in the forthcoming section.

4.3.1. Activation increases in prefrontal regions at capacity limitations only in high

performers? Experiment 2

As stated before, the results of my previous study (Jaeggi et al., 2003) were the basis for the following experiment, which was conducted in the project No.15-640, funded by the Swiss National Science Foundation (SNF). Here I was mainly interested in the functional processes taking place in prefrontal cortices at capacity limitations, and to that end, I used the dual-task paradigm at various levels of load. Based on the findings as reported in Jaeggi et al. (2003), I predicted an increase in cortical activation with increasing processing demand mainly in prefrontal regions. A decrease in cortical activation in the dual-task at high levels of load was only expected, if the participants were not able to perform the required task adequately anymore, which was assessed by behavioural data (accuracy and reaction times). Therefore, I predicted a differential activation pattern for high- and low-performing participants, i.e., a decrease in activation in the most difficult dual-task condition for the low-performing individuals compared to an ongoing increase in the high-performing individuals.

4.3.2. Method

Participants. 15 young adults (9 women, 6 men; age range 21 to 29 years; mean age 25.3) participated in this study. All were assessed as right-handed (Oldfield, 1971) with no seeing, hearing or neurological disorders, and normal structured MRI scans. All subjects were given identical practice with the task before scanning[6]. Based on their performance levels in the most difficult task in the fMRI-session (see below), three groups of participants were selected and consisted of 4 low performing-subjects (2 men, 2 women), 4 high-performing subjects (4 women) and 7 participants perform-ing at intermediate levels (4 men, 3 women).

Stimulus Material and Procedure. The sequential n-back paradigm was used again, however, instead of the shapes administered in Jaeggi et al. (2003), we used visuospatial-nonverbal[7] and auditory-verbal material, also administered as dual-task as described in Experiment 1 (see also Figure 2). We used a blocked periodic design in which the stimuli were presented to the participants in 30s (i.e., 0-back) and 60s (i.e., 1-back to 3-back) epochs, always preceded by instruction templates for 15s (see Figure 5). The positions of the 1- to 3-back tasks were determined randomly, each appearing twice and always followed and preceded by the 0-back task, which served as baseline task. All conditions were matched for number of targets presented (33%) and consisted of 70 trials in the 0-back task, and of 40 trials in the 1- to 3-back tasks; one trial consisting of one stimulus (500 ms) and one ISI (2500 ms). Targets and non-targets were determined pseudorandomly, i.e., varying the position of the targets in each sequence randomly, while holding the number of distracters constant (e.g. 2-back targets in a 3-back condition). Participants responded by pressing a pneumatic squeeze ball (left hand for visual stimuli; right-hand for auditory-verbal stimuli), if the current stimulus was the same as the one presented n back in the sequence. With a post-test inventory we assessed the used strategies as well as the subjectively experienced difficulty-level for each task.

---

[6] All the tasks were practiced in a similar session as in the fMRI-session in the experimental laboratory beforehand. However, the sequence of the different n-back tasks varied, being randomly assigned by the program in order to prevent position-effects.

[7] The stimuli were changed, because experimental evidence with the visuospatial stimulus material indicated less verbalization processes and therefore a probably lesser overlap of resources in the dual-task conditions, expressed behaviourally and on neuronal level. This is also supported by the literature (e.g. Carlson et al., 1998).
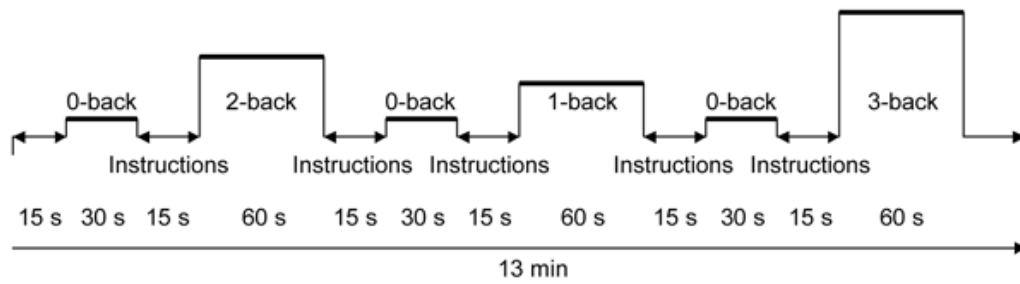
Figure 5. Task alternations and instructions for the n-back task as used in Experiment 2 and 4.

fMRI methods. After shimming, whole brain and high resolution fMRI was performed with a blood oxygenation level dependent (BOLD) echo-planar imaging (EPI) sequence (matrix 128 x 128 x 48 slices x 108 measurements, resulting in 1.5 x 1.5 x 3 mm x TR 7500 ms; TE 60 ms) on a standard clinical 1.5-T whole-body MRI scanner (Magnetom Vision, Siemens Medical Systems, Erlangen, Germany). Slice angulation was parallel to a line connecting the base of the genu corporis callosi and the confluens sinuum, resulting in an angle of about 20–30° with respect to the bi-commissural (AC–PC) line. Additionally, standardized sets of high-resolution structural images were acquired for later coregistration. The experiment was controlled by a Microsoft Windows based personal computer using E-Prime (Psychology Software Tools, Pittsburgh, PA). Stimuli were visually presented to the subjects via a prismatic mirror and aurally presented via headphones as they were laying in the scanner.

Data Analysis

Behavioural Data. For behavioural data, RTs and accuracy (i.e., P_r; hits minus false alarms according to Snodgrass & Corwin (1988)) were analyzed, and additionally, groups of high-, low-, and intermediate-performing participants were differentiated post-hoc with a frequency analysis based on their accuracy levels in the most difficult task (3-back), defining the individuals performing in the lowest quartile as low performer, those in the highest quartile as high performer, and those in-between as intermediate performer. A 2-way repeated measures ANOVA was calculated for both behavioural measures (accuracy and RT) with load (0 to 3-back) as within-factor, and performance group as between-factor and post-hoc tests were applied in case of a significant interaction.

fRMI Data. For fMRI evaluation, *z* score maps were generated after spatial filtering and projected onto original EPI images, using self-developed software, which is described somewhere else (Jaeggi et al., 2003; Nirkko, 2000). To compare the amount of activation in a given area across experimental conditions, a priori Volumes of Interest (VOIs) were defined using the individual anatomical parcellation method according to Rademacher et al. (1992). The same VOIs were defined for each participant, thus preserving individual differences in brain structure. The VOIs were manually drawn on each functional slice, also with self-developed software (Jaeggi et al., 2003). I defined 34 VOIs in each hemisphere. The main focus of interest was on the DLPFC, which was defined as the middle frontal gyrus (F2; BA 6, 8, 9, and 46) with the exception of the posterior-most part (frontal eye fields) to exclude activation related to eye movements. The interrater reliability of this VOI-defining procedure between two staff members was evaluated in our previous study (Jaeggi et al., 2003) and ranged between $r = 0.82$ and $r = 0.93$, reflecting a high degree of reliability.

For each VOI, 3-way repeated-measurements ANOVAs were calculated with task load (1 to 3-back) and hemisphere (left vs. right) as within-factors, and performance group (low, intermediate and high performer) as between-factor. Post-hoc tests were administered in case of significant interactions.

4.3.3. Results

Behavioural Data

Generally, participants performed very well in all tasks. Even at high levels of load, several participants were still able to perform the task adequately, but, as predicted, accuracy levels were near chance level in the most difficult condition (3-back dual task) in some participants. A post-hoc frequency analysis based on the performance in the most difficult task (3-back dual; accuracy) resulted in 4 low-performing participants ($P_r < 0.36$; $M = 0.29$; $SD = 0.08$) and 4 high-performing participants ($P_r > 0.66$; $M = 0.71$; $SD = 0.04$), with 7 participants being at intermediate levels ($M = 0.50$; $SD = 0.05$).

The 3-way repeated measures ANOVA indicated a significant main effect of load, i.e., 0-back to 3-back tasks (*accuracy*: $F(1.96,14) = 201.76$; $p < 0.001$; *RT*: $F(1.52,14) = 23.79$; $p < 0.001$) and a significant main effect for performance group in accuracy ($F(2,14) = 11.03$; $p < 0.01$). The interaction between load and performance group was significant for accuracy only ($F(6,7) = 11.53$, $p < 0.001$). Post hoc tests for accuracy yielded significant differences between the three performance groups at the 3-back level only (*high vs. low*: $t(6) = 9.40$; $p < 0.001$; *high vs. intermediate*: $t(9) = -7.30$; $p < 0.001$; *intermediate vs. low*: $t(9) = 5.21$; $p = 0.001$). Descriptive data for both behavioural measures are plotted on Figure 6.

The qualitative analysis of the post-test inventory revealed that most of the participants used a visual tracking strategy for the visuospatial task and a verbal rehearsal strategy for the auditory task, i.e., combining the two strategies as it was a dual-task situation. However, all our high-performing participants mentioned fewer explicit, if any strategies at all. Basically they stated not to rely on strategies, but rather experienced a kind of intuitive problem solving whereas the low-performing participants combined and tried out different strategies, changing them frequently during task execution.
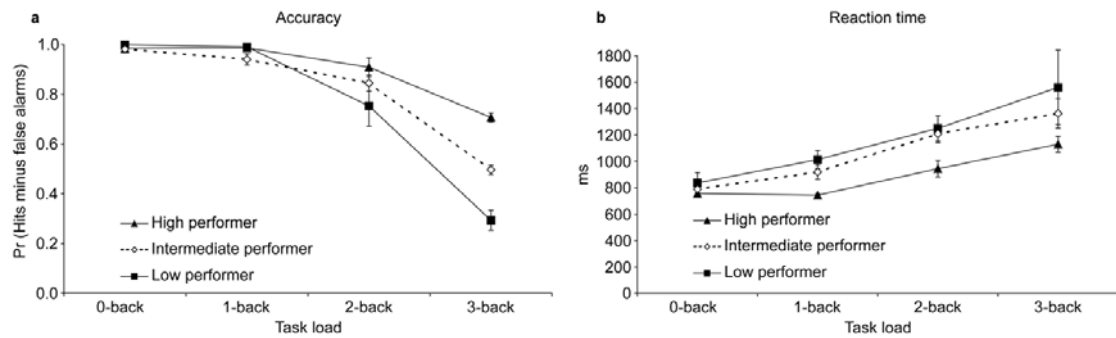
Figure 6. Means and standard errors of the mean (SEM) for the three groups of participants for each level of load. (a) Accuracy. (b) Reaction times (hits only).

Imaging Data

The activation data were evaluated for all 34 VOIs and activation changes in respect to the different levels of load compared to the baseline are shown on Table 7, separately for each performance group and hemisphere. The activation changes in the 3-back task for the high- and low-performing participants are additionally illustrated on Figure 7.

Table 7. Activation changes in all defined volumes of interest (VOIs) for each level of load (compared to the baseline task, i.e., 0-back) and for each hemisphere, reported for all three groups of participants.

Brain Area	Parcellation Units (VOIs)	Abbr.	Left Hemisphere									Right Hemisphere								
			Low Performer			Intermediate			High Performer			Low Performer			Intermediate			High Performer		
			1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back
Lateral cerebral surface																				
Frontal lobe	Central sulcus	ce	-0.09	0.18*	-0.14*	-0.01	-0.03	0.05	0.08	0.18**	-0.12*	-0.02	0.09	-0.08	-0.03	-0.05	-0.15***	0.00	0.10*	0.09*
	Superior frontal gyrus	F1	0.02	0.57**	-0.02	0.03	0.03	0.15***	0.10	0.05	-0.32**	0.04	0.33**	0.18*	0.02	0.14**	0.26***	0.09	0.05	-0.07
	Middle frontal gyrus (DLPFC)	F2	0.02	0.44***	0.44***	0.10**	0.34***	0.41***	0.12*	0.32***	-0.04	0.16**	0.22**	0.31***	0.12***	0.16**	0.26***	0.04	0.19***	-0.11
	Frontal eye fields	FEF	0.18**	0.31**	0.25***	0.13**	0.34***	0.25***	0.19*	0.32***	0.27**	0.10	0.18*	0.12*	0.09	0.31***	0.07	0.25**	0.23**	0.16*
	Frontal pole	FP	-0.23**	0.01	0.26***	-0.16**	-0.14**	0.18**	0.14	0.04	-0.24	-0.25*	-0.12	0.29**	-0.20**	0.19**	0.05	0.10	0.13	0.03
	Inferior frontal sulcus	if	0.18**	0.41***	0.38***	-0.02	0.25**	0.35***	0.04	0.19**	-0.26**	0.24***	0.25***	0.31***	0.07*	0.19***	0.16***	0.07	0.11**	-0.09
	Precentral sulcus[1]	prc	0.12*	0.55***	0.29***	0.20***	0.45***	0.57***	0.22**	0.42***	0.07	0.15*	0.35***	0.29**	0.18***	0.39***	0.36***	0.01	0.19***	-0.02
	Superior frontal sulcus	sf	0.20**	0.66***	0.27***	0.11***	0.36***	0.48***	0.15	0.24***	-0.17	0.02	0.16**	0.32***	0.10***	0.33***	0.31***	-0.12	0.21***	-0.24**
Parietal lobe	Inferior intraparietal lobule[2]	INFP	0.08	0.48***	0.24**	0.01	-0.01	0.15**	0.23***	0.48***	0.13*	0.15**	0.29***	0.28***	0.05	0.21***	0.19***	0.06	0.14***	0.04
	Intraparietal sulcus	ip	0.34***	0.66***	0.45***	0.16***	0.41***	0.58***	0.30***	0.56***	0.23***	0.31***	0.57***	0.49***	0.16***	0.35***	0.44***	0.26***	0.37***	0.29***
	Postcentral sulcus	poc	-0.07	0.45*	-0.18	-0.01	0.13**	0.05	-0.05	0.17*	0.00	0.01	0.00	-0.05	0.04	0.16***	0.02	-0.03	0.03	-0.10
Temporal lobe	Inferior temporal sulcus	it	0.25*	0.19*	0.16*	-0.02	0.08	0.05	0.32*	0.14**	-0.41***	0.14*	0.05	0.00	0.12***	0.19***	0.15***	0.26	-0.08	-0.31***
	Superior temporal sulcus	st	0.08	0.18***	0.15**	-0.02	0.11***	0.09**	0.25*	0.18***	-0.35***	0.13*	0.16***	0.06	0.05*	0.22***	0.09*	0.04	0.01	-0.17***
	Temporal pole	TP	0.50*	0.31*	-0.18	-0.15**	-0.07	-0.02	0.09	-0.21*	-0.49***	0.33**	0.14	-0.13	-0.08	0.11	-0.23	-0.02	-0.19	0.00
Occipital lobe	Intracalcarine cortex	CALC	0.13*	0.23***	0.09*	0.10*	0.30***	0.15***	0.12	0.16***	-0.07	0.16*	0.24***	0.23***	0.05	0.30***	0.09*	0.22**	0.13**	-0.13***
	Lateral occipital cortex, inferior division	OLi	0.44***	0.30***	0.05	0.12**	0.14*	0.12*	0.28*	0.28***	-0.11	0.31***	0.16**	-0.06	0.13***	0.22***	0.15***	0.24*	0.09*	-0.05
	Lateral occipital cortex; superior division	OLs	0.44***	0.61***	0.28***	0.21***	0.47***	0.45***	0.32***	0.59***	0.31***	0.35***	0.47***	0.23***	0.18***	0.44***	0.45***	0.22**	0.38***	0.25***
	Occipital pole	OP	0.55**	0.26**	-0.20	0.19*	0.27***	0.05	0.71*	0.09	-0.53**	0.57***	0.44***	0.20*	0.23***	0.35***	-0.04	0.69*	0.26*	-0.20
Medial cerebral surface																				
Frontal lobe	Cingulate gyrus; anterior division	CGa	-0.08	0.03	0.04	-0.09**	0.07*	0.08	-0.01	-0.03	-0.26***	0.01	0.02	-0.02	0.10**	0.11	0.08	-0.08	0.15**	0.09
	Cingulate Gyrus, posterior division	CGp	-0.11	0.01	0.10*	-0.07*	0.11**	0.00	0.11	0.04	-0.22**	0.01	0.09	0.21**	0.03	0.14**	0.01	0.17	-0.13*	-0.36***
	Marginal ramus of cingulate sulcus	ma	-0.03	0.07	0.07	0.09**	0.21***	0.14**	0.20**	0.09	0.08	0.02	0.17	0.06	0.08	0.35***	0.25***	0.13	0.03	0.01
	Paracingulate gyrus	PAC	-0.06	0.33***	0.31*	-0.06	0.19***	0.36***	0.16**	0.09	-0.14*	-0.07	0.20**	0.52***	0.26***	0.41***	0.37***	-0.06	0.38***	0.37***
	Supplementary motor cortex	SMC	-0.13	0.11	0.00	-0.11**	0.13**	0.09	0.11	0.10	-0.19**	-0.24*	0.05	0.27*	0.09*	0.15	-0.16*	-0.12	0.09	0.46***

Table 7. (Continued).

Brain Area	Parcellation Units (VOIs)	Abbr.	Left Hemisphere									Right Hemisphere								
			Low Performer			Intermediate			High Performer			Low performer			Intermediate			High Performer		
			1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back	1-back	2-back	3-back
Medial cerebral surface																				
Parietal lobe	Parietooccipital sulcus	po	0.15*	0.33***	0.24***	0.11**	0.37***	0.45***	0.15**	0.52***	0.29***	0.20**	0.45***	0.30***	0.11**	0.37***	0.34***	0.17**	0.33***	0.17*
	Subparietal sulcus	sp	-0.11	0.13**	0.19***	-0.19***	-0.07	-0.14**	0.04	0.09	-0.06	-0.02	0.24***	0.19**	-0.07	0.02	-0.09	0.12	-0.05	-0.16**
Intrasylvian cerebral surface	Heschl sulcus	he	0.06	-0.12	-0.03	-0.12*	0.06	0.14*	0.12	-0.25**	-0.67***	-0.23*	-0.44***	-0.44***	0.11	-0.18	-0.25*	-0.24	-0.31***	-0.33***
	Insular cortex	INS	0.00	0.05	-0.14**	0.02	0.11***	0.09**	-0.06	0.03	-0.09	0.00	0.11	0.00	-0.01	0.10***	0.08**	-0.05	0.05	-0.10*
	Sylvian fissure (total extent)	sy	-0.13	-0.13	-0.31**	-0.04	0.17***	0.15**	-0.06	-0.12*	-0.48***	-0.09	-0.13	-0.34***	0.02	0.00	-0.30***	-0.15	-0.17**	-0.25**
Subcortical structures	Caudate nucleus	cau	0.10	0.22*	0.17*	-0.01	0.20***	0.32***	-0.03	0.16	-0.36***	0.26**	0.09	0.22*	0.26***	0.22	0.03	-0.12	0.26**	-0.01
	Putamen	pu	0.13*	0.02	-0.01	0.06	0.08	0.03	0.21**	0.13*	-0.11	0.04	0.10	-0.12	0.01	0.18***	0.03	-0.09	0.09	-0.27**
	Thalamus	thal	0.11	0.13*	0.16***	0.04	0.18***	0.13***	0.13*	0.22***	-0.09	0.15**	0.14**	0.10	0.09**	0.26***	0.11**	0.05	0.14**	-0.05
Cerebellum	Cerebellar hemisphere	CH	0.23**	0.29***	0.22***	0.08**	0.24***	0.13**	0.34**	0.28***	0.15*	0.17*	0.30***	0.17*	0.07*	0.23***	0.10*	0.18*	0.36***	0.09
	Cerebellar vermis	CV	0.01	0.11	0.28***	0.13***	0.20***	0.12*	0.10	0.12**	0.04	0.11*	0.07	0.21***	0.06*	0.18***	0.18***	0.17*	0.17***	0.10*

Note: Values are given as group average of percent BOLD signal change (positive values = activation; negative values = deactivation) with statistical group analysis of fMRI data from VOIs. p values as two-tailed significance levels set at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
[1] Premotor cortex adjacent to the precentral sulcus at the level of the "hand notch".
[2] Inferior parietal lobule comprising the intermediate sulcus of Jensen, the angular sulcus, and the anterior occipital sulcus.
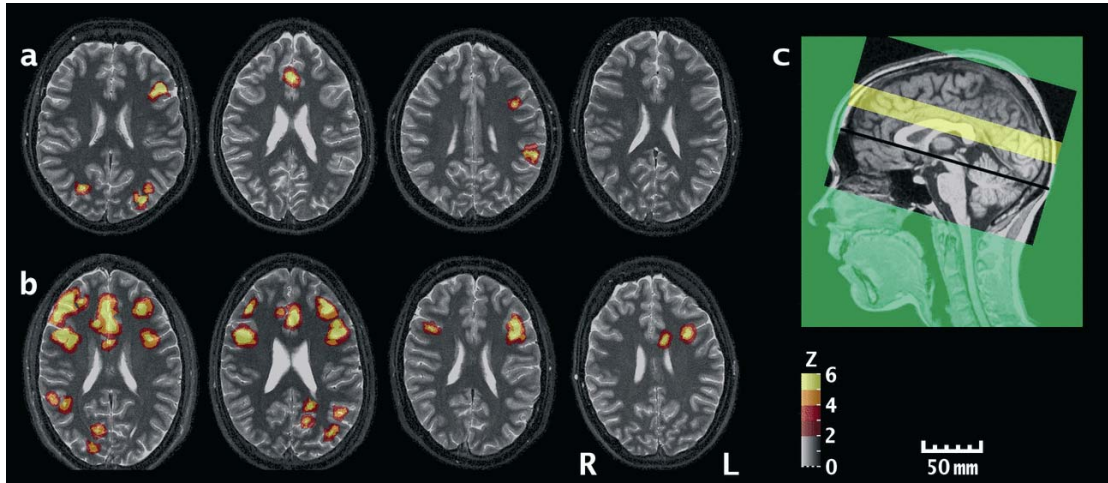
Figure 7. Activation of the 3-back vs. 0-back contrast, shown for each of the 8 high- and low-performing participants: (a) High performer. (b) Low performer. (c) Covered fMRI volume with angulation reference (Nirkko et al., 2001) (black line). To show the *z* score maps of a representative part of each brain on just one panel each, a maximum intensity projection of 5 adjacent functional slices (yellow band) was overlaid onto one structural slice. Clusters containing a *z* score of at least 5.0 ($p < 0.05$, Bonferroni corrected for multiple comparisons) were included, with their extent colour coded down to a *z* score of 2.0 ($p < 0.05$, uncorrected). Distance scale: 50 mm.

Since I was primarily interested in the role of the prefrontal cortex in capacity limitations and also for reasons of shortness, the main focus in the following section will lie on the data regarding the DLPFC, and therefore, the ANOVAs reported here do only concern the DLPFC. The 3-way repeated-measurements ANOVA with load (1-back to 3-back) and hemisphere (left vs. right) as within-subject factors, and performance group (low, intermediate and high performer) as between-subject variable yielded a main effect of load ($F(2, 1259) = 21.23$; $p < 0.001$; *linear trend*: $F(2, 1259) = 23.28$; $p < 0.001$), a main effect for hemisphere ($F(1, 1259) = 14.25$; $p < 0.001$), and a main effect for performance group ($F(2, 1259) = 19.70$; $p < 0.001$). There were two significant 2-way interactions, namely load x hemisphere ($F(2, 1259) = 6.60$; $p = 0.001$), and load x performance group ($F(4, 1259) = 14.27$; $p < 0.001$), which are shown on Figures 8 and 9. The interaction hemisphere x performance group, as well as the 3-way interaction was not significant. Post-hoc tests resulted in significant differences between hemispheres in the 2- and 3-back tasks (*2-back*: $t(418) = 4.32$; $p < 0.001$; *3-back*: $t(418) = 2.69$; $p < 0.01$; two-tailed) with higher activations in the left hemisphere. In the 1-back task, no difference was observed between hemispheres (see Figure 8). Load had a differential impact on the activation changes in the different performance groups, corresponding to the accuracy data: Differences between groups were observed in the in the most difficult condition, i.e., in the 3-back

task only, where the high-performing group had significantly lower activations than the low- ($t(222) = 7.28$; $p < 0.001$) and the intermediate-performing group ($t(194.02) = 7.49$; $p < 0.001$; both two-tailed). There were no differences between the low- and the intermediate-performing group; and no differences between groups were observed at the other levels of load. On Figure 9, activation patterns for the 3-back task are

Activation changes in the DLPFC
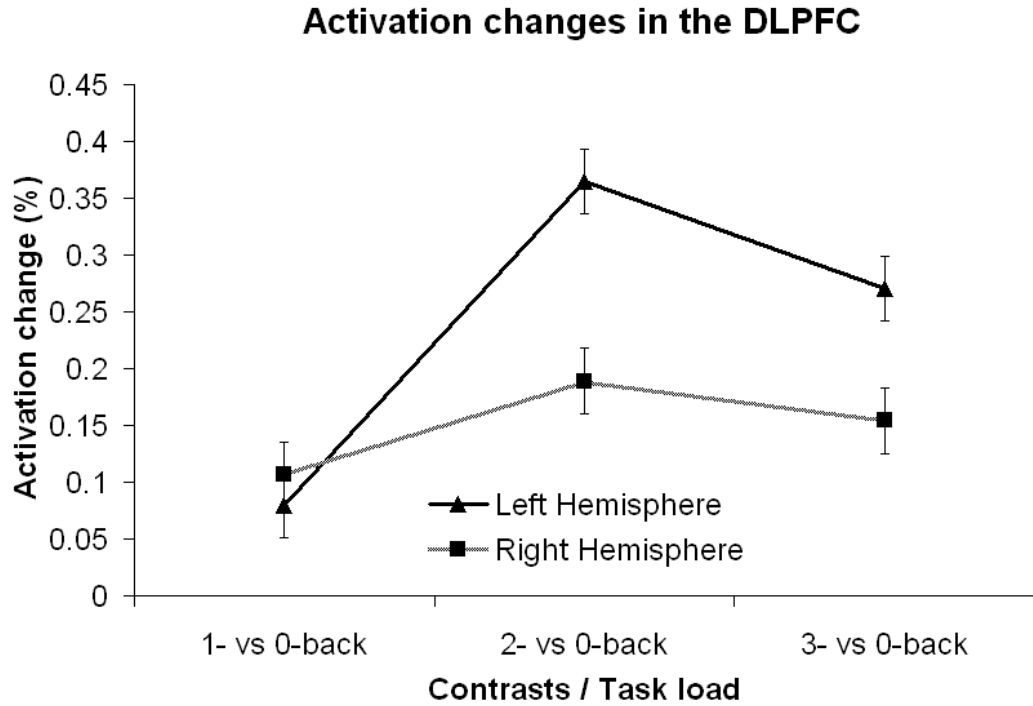
shown for the high- and the low-performing participants.

Figure 8. Hemisphere x load. Mean BOLD signal changes (in %) and SEM in the DLPFC, as difference between each load condition and the baseline-condition (0-back), shown for the two hemispheres.

As can be seen on Table 8, there are only a few VOIs which do not indicate significant interactions for load x performance group (i.e., sy, poc, FEF, CH, and CGa), however, it is important to note, that concerning the 3-back task, the high-performing participants do not seem to compensate their absent activation in the DLPFC in recruiting different areas. Rather, they consistently show deactivations in most of the VOIs. The only VOI showing higher activation than any other performance group is the right supplementary motor cortex (0.46 vs. 0.27 in the low-, and -0.16 in the intermediate-performing group).
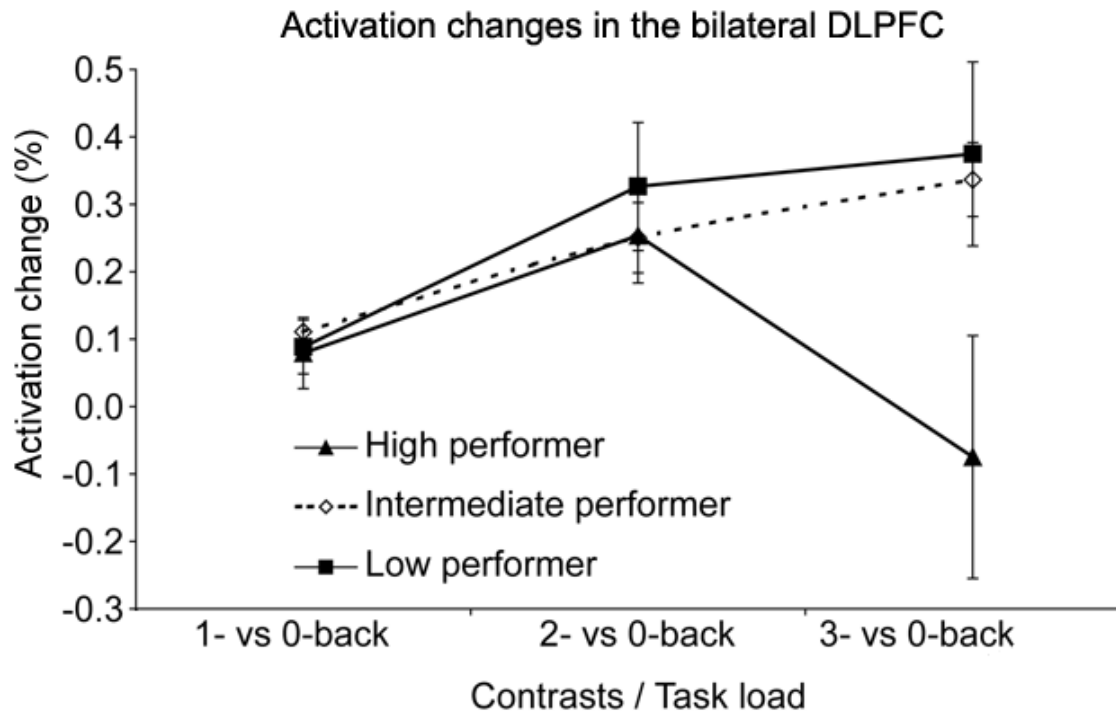
Figure 9. Load x performance group. Mean BOLD signal changes (in %) and SEM in the bilateral DLPFC, as difference between each load condition and the baseline-condition (0-back), shown for the three performance groups.

Table 8. Main effects and interactions of the 3-way ANOVA shown for each defined VOI (alphabetically).

VOI	Load			Hemisphere			Performance group			Load x Hemisphere			Load x Performance group			Hemisphere x Performance group			Load x Hemisphere x Performance group		
	F	df	p	F	df	p	F	df	p	F	df	p	F	df	p	F	df	p	F	df	p
CALC	18.52	2	0.000	0.21	1	0.644	8.39	2	0.000	0.21	2	0.807	7.82	4	0.000	1.60	2	0.203	1.25	4	0.289
CGa	3.42	2	0.033	8.77	1	0.003	3.93	2	0.020	0.15	2	0.861	1.00	4	0.409	2.28	2	0.103	4.61	4	0.001
CGp	3.51	2	0.030	0.72	1	0.398	6.44	2	0.002	1.65	2	0.193	15.00	4	0.000	3.20	2	0.041	0.47	4	0.758
CH	10.85	2	0.000	1.61	1	0.204	6.31	2	0.002	1.42	2	0.242	1.56	4	0.183	0.11	2	0.892	0.60	4	0.661
CV	2.67	2	0.070	0.58	1	0.448	0.48	2	0.616	0.20	2	0.823	4.51	4	0.001	1.00	2	0.369	1.52	4	0.194
F1	9.83	2	0.000	2.04	1	0.153	11.22	2	0.000	4.59	2	0.010	12.96	4	0.000	0.58	2	0.563	1.65	4	0.160
F2	21.23	2	0.000	14.25	1	0.000	19.70	2	0.000	6.60	2	0.001	14.27	4	0.000	0.23	2	0.798	1.41	4	0.229
FEF	6.09	2	0.002	6.59	1	0.010	0.74	2	0.475	1.22	2	0.296	1.01	4	0.399	0.34	2	0.711	0.31	4	0.875
FP	7.39	2	0.001	0.99	1	0.321	0.42	2	0.659	0.86	2	0.423	8.37	4	0.000	0.82	2	0.439	2.74	4	0.028
INFP	14.46	2	0.000	2.40	1	0.122	12.74	2	0.000	1.46	2	0.233	5.27	4	0.000	12.38	2	0.000	3.60	4	0.006
INS	11.70	2	0.000	1.24	1	0.265	11.68	2	0.000	0.49	2	0.610	2.95	4	0.019	1.91	2	0.148	0.46	4	0.762
OLi	21.20	2	0.000	2.41	1	0.121	2.02	2	0.134	0.56	2	0.572	6.47	4	0.000	2.49	2	0.083	0.77	4	0.542
OLs	20.46	2	0.000	7.48	1	0.006	0.95	2	0.386	1.00	2	0.369	7.00	4	0.000	1.43	2	0.240	0.18	4	0.951
OP	36.31	2	0.000	4.10	1	0.043	1.96	2	0.141	0.99	2	0.372	6.92	4	0.000	1.10	2	0.333	0.88	4	0.475
PAC	39.02	2	0.000	23.29	1	0.000	7.42	2	0.001	5.17	2	0.006	5.94	4	0.000	3.97	2	0.019	12.96	4	0.000
SMC	8.55	2	0.000	2.32	1	0.128	0.97	2	0.381	5.62	2	0.004	3.24	4	0.012	1.51	2	0.222	11.50	4	0.000
TP	9.31	2	0.000	0.07	1	0.784	9.68	2	0.000	0.83	2	0.436	3.83	4	0.004	1.22	2	0.295	2.68	4	0.030
cau	3.79	2	0.023	1.36	1	0.243	8.01	2	0.000	0.67	2	0.512	3.07	4	0.016	0.73	2	0.483	4.55	4	0.001
ce	13.31	2	0.000	0.46	1	0.497	6.22	2	0.002	1.21	2	0.297	4.24	4	0.002	2.71	2	0.067	5.39	4	0.000
he	8.49	2	0.000	14.14	1	0.000	12.18	2	0.000	0.21	2	0.814	2.68	4	0.030	3.84	2	0.022	6.97	4	0.000
if	7.63	2	0.001	0.53	1	0.469	28.33	2	0.000	2.65	2	0.071	9.33	4	0.000	1.06	2	0.346	1.94	4	0.101
ip	26.00	2	0.000	3.52	1	0.061	9.15	2	0.000	1.34	2	0.262	8.58	4	0.000	0.28	2	0.758	1.47	4	0.208
it	16.58	2	0.000	0.67	1	0.414	5.90	2	0.003	0.71	2	0.493	13.09	4	0.000	5.72	2	0.003	0.85	4	0.495
ma	2.16	2	0.116	0.53	1	0.465	8.34	2	0.000	0.57	2	0.566	3.87	4	0.004	2.40	2	0.091	0.41	4	0.801
po	28.47	2	0.000	0.59	1	0.442	0.25	2	0.776	0.77	2	0.462	3.19	4	0.013	3.11	2	0.045	0.81	4	0.519
poc	17.04	2	0.000	2.26	1	0.133	1.47	2	0.229	5.51	2	0.004	1.72	4	0.144	1.22	2	0.297	4.08	4	0.003
prc	35.17	2	0.000	21.59	1	0.000	28.73	2	0.000	1.37	2	0.253	9.23	4	0.000	1.74	2	0.176	2.68	4	0.030
pu	13.95	2	0.000	5.55	1	0.019	2.48	2	0.084	3.94	2	0.020	3.05	4	0.016	3.74	2	0.024	0.25	4	0.910
sf	27.03	2	0.000	22.54	1	0.000	37.10	2	0.000	1.65	2	0.192	13.34	4	0.000	2.43	2	0.088	6.05	4	0.000
sp	4.99	2	0.007	1.59	1	0.207	19.87	2	0.000	1.41	2	0.245	7.13	4	0.000	2.78	2	0.062	0.61	4	0.659
st	16.69	2	0.000	0.14	1	0.707	9.91	2	0.000	0.68	2	0.508	13.57	4	0.000	2.92	2	0.054	4.68	4	0.001
sy	16.30	2	0.000	3.02	1	0.083	21.57	2	0.000	0.88	2	0.415	1.58	4	0.176	5.89	2	0.003	5.88	4	0.000
thal	12.32	2	0.000	0.05	1	0.821	4.48	2	0.012	0.10	2	0.903	5.77	4	0.000	1.35	2	0.260	1.44	4	0.219

Note: *N* =1260 (number of measurements); see Table 7 for abbreviations.

4.3.4. Discussion

In accordance with the results of our prior study (Jaeggi et al., 2003), an increase in prefrontal activation was observed with increasing load, as also expressed with the significant linear trend. More activation changes were observed in the left hemisphere, especially at higher levels of load. This is in accordance with findings in the literature (e.g. Mitchell, Johnson, Raye, & Greene, 2004; Ranganath, Johnson, & D'Esposito, 2000), suggesting that left lateralized activation in the prefrontal cortex (PFC) occurs, when monitoring or evaluative processes are engaged in retrieval processes, which are essential for specifying the source of a memory (Johnson, Hashtroudi, & Lindsay, 1993). Therefore, the left PFC seems to be recruited specifically during memory tasks that demand specific source attributions; processes which seem critical in an n-back task at higher levels of load. It has been further suggested, that the left PFC is engaged in processing sequential dependencies between single adjacent events (Wood & Grafman, 2003). There was no interaction between hemisphere and performance group, thus indicating, that this left lateralization primarily reflected the above-mentioned task-related requirements and not being related to individual differences in task performance.

As predicted, as cognitive processing load increases, task performance decreases. Additionally, a differential activation pattern was observed between the three performance groups, but only on the highest level of load (3-back dual task), which most certainly reflected strategic differences. In contrast to our predictions however, the low-performing subjects showed a steeper activation increase from 1-back to 3-back, which can be interpreted in terms of additional processes necessary to perform the task, such as attentional processes and mental effort. A decrease in activation was observed in the high-performing participants from the 2- to 3-back task, suggesting a change to a very efficient use of strategies in this group, which is also reflected in the behavioural data: The performance in this group was well beyond all other participants' and also far from chance level in the most difficult task. To illustrate this exceptional performance, in Experiment 1 (N =132), only two participants were able to reach the high performer's performance level of more than 0.66 (P_r).

Most striking in our study was now the observation, that there was a completely different activation pattern in the forebrain of low performing-participants compared to high performers: Low-performing participants showed an activation-increase with increasing processing load, although their performance was near chance

level in the most difficult condition. Similarly, the averaged data over all subjects showed load-dependent activation, a finding that is consistent with many studies (e.g. Owen et al., 2005; Rypma & D'Esposito, 1999; Wager & Smith, 2003) and also with our previous study (Jaeggi et al., 2003). The question therefore concerns the activation decrease in the high-performing participants, which is not only expressed in the DLPFC, but also in many other VOIs. Although, there have been reports of a lower activation in the DLPFC in high-performing participants (Rypma & D'Esposito, 1999; Rypma & D'Esposito, 2000), it has also been shown, that high-performing participants nevertheless showed an activation increase with increasing processing demand (Rypma et al., 2002), which was clearly not the case in our individuals. The differential activation pattern of good and bad performers of the present study clearly shows that brain activation, cognitive load and performance are not related in a simple and linear way. Activation clearly discriminates high- from low-performing individuals, but nevertheless seems only to be related indirectly to performance. The high brain activity of low performers in the extremest mental load condition is either not relevant or impairs performance. On the other hand, the brain of good performers 'shuts down' in terms of brain activity in conditions of extreme cognitive demand. There seems to be no compensating brain area for these missing activations, apart from the right SMC, but it seems nevertheless very unlikely, that one single area would be responsible for the expertise expressed in the high-performing participants. It seems rather probable, that more effective functional connectivity might account this expertise; an fMRI-data analysis approach put forward for example by McIntosh (1999) going beyond the univariate approach used here, determining effective connectivity by means of structural equation modeling.

Nevertheless, it can be argued with the behavioural data, that the high-performing participants have not yet reached their capacity-limits, and that no activation increase was observed for this reason. It would be interesting to see their activation pattern, when they are confronted with even more difficult conditions (e.g. 4- and 5-back tasks); an experiment which will be discussed in the next chapter.

4.4. Interindividual differences beyond capacity limitations

4.4.1. Does the brain of high-performing participants 'stay cool' even at their

capacity limit? Experiment 3

As described in Experiment 2, there are large interindividual differences in the performance of the dual n-back task at high levels of load which is also expressed on neuronal level. A striking result was the observation that the high-performing group did not seem to have reached their capacity limits in the most difficult 3-back condition. The question therefore was, whether this group would also show activation increases in prefrontal areas, when being challenged to their capacity limits: Would their brain react similarly as the brain of low-performing participants, only at later stages, i.e., showing an activation increase at higher levels of load? Or would their brains 'stay cool' all the same, irrespective of the load placed on processing capacity? Similarly, what would happen at such stages of cognitive overload in low-performing participants, being at their limit already at the 3-back level? Would we be able to observe a decrease in activation at the highest load, as proposed by D'Esposito (2001)? To answer these questions, I recruited all participants falling into the two high- and low-performing categories to participate in a follow-up experiment. This experiment consisted of the same dual-task experiment as before, and additionally, a block with 4- and 5-back dual-tasks was administered, which I assumed would challenge every participant beyond the limits. Thus, I wanted to replicate the results obtained in Experiment 2, and further, I examined possible differential activation patterns in these two groups of participants, challenging the proposed capacity-limits to an extent, where performance most likely drops to chance-level in every performer.

4.4.2. Method

Participants. 8 young healthy and right-handed adults (6 women, 2 men; age range 21 to 29 years; mean age 24.4) participated in this study. They were selected on the basis of their performance-levels in Experiment 2 and scanned approximately 6 month after the first session (Experiment 2). The participants were all given identical practice with the task just before the scanning took place: Again the n-back task was explained to them, and they were allowed to make a few practice trials, until they felt confident, that they understood the tasks.

Material and Procedure. The sequential n-back paradigm with visuospatial-nonverbal and auditory-verbal material was performed as dual-task version at six levels of difficulty (0- to 5-back) in a blocked periodic design. First, participants performed a 13-minute-lasting block with 0- to 3-back tasks, exactly as in Experiment 2 in order to replicate the previously obtained results, alternating between the baseline task (0-back) and a randomly assigned n-back task. After that, high-resolution MRI was performed as described in Experiment 2. As last block, 4- and 5-back tasks were administered, also alternating randomly with 0-back tasks, again serving as baseline.

Analysis. Behavioural data as well as functional data were analyzed as described in Experiment 2. Further, the behavioural data and the activation patterns in the DLPFC in sessions 1 (Experiment 2) and 2 (Experiment 3) were compared in order to compare performance and the BOLD-response between sessions with a repeated measures ANOVA with load (1 to 3-back) and session as within-subject variables, and with performance group as between-subject variable for the behavioural date, i.e. with load, hemisphere, session, and performance group for the functional data, respectively. As additional analysis, correlations of the activations between the two sessions were calculated for the DLPFC in order to estimate test-retest reliability of the BOLD-response. Further, correlations between the BOLD-response and the performance measures were calculated in order to obtain an estimate of the relationship between brain activity and behaviour. For these measures, non-parametric bootstrapping with $R = 9'999$ replicates of Fisher's z transformation at the 95% confidence interval were applied in order to test the null hypothesis of zero correlation.

4.4.3. Results

Behavioural Data. Again, participants performed surprisingly well in this challenging dual task. Even at highest levels of load, several participants were still able to perform the task adequately, but, as predicted, the accuracy levels were approaching chance level in the most difficult condition (5-back task) in most participants and there were no differences between groups anymore. The repeated measures ANOVA indicated a main effect of load (*accuracy*: $F(5,7) = 157.18$; $p < 0.001$; *RTs*: $F(5,7) = 11.32$; $p < 0.001$), a significant main effect for performance group for accuracy ($F(1,7) = 6.71$; $p < 0.05$), and also significant interaction between load and performance group (*accuracy*: $F(5,7) = 2.83$; $p < 0.05$;). Post-hoc tests for accuracy revealed significant group differences at the 3-back level again as in Experiment 2, with the high-performing group scoring significantly higher ($t(6) = 3.17$; $p < 0.05$; one-tailed). Other group-differences were observed in the 1-back ($t(6) = 2.44$; $p < 0.05$) and in the 4-back task ($t(6) = 1.97$; $p < 0.05$; both one-tailed), again the high-performing group being better (see Figure 10). In order to reveal any differences between sessions, a 3-way repeated measures ANOVA was calculated for accuracy with load (1 to 3-back) and session (Experiment 2 vs. Experiment 3) as within-factor, and performance group as between-factor was calculated. There was a significant main effect for load ($F(1.76, 7) = 71.76$; $p < 0.001$), a significant main effect for session ($F(1, 7) = 78.51$; $p < 0.001$) and a main effect for performance group ($F(1, 7) = 15.69$; $p < 0.01$). All 2-way interactions were significant (*load x performance group*: $F(3, 7) = 11.02$; $p < 0.001$; *session x performance group*: $F(1, 7) = 8.78$; $p < 0.05$; *session x load*: $F(3, 7) = 5.65$; $p < 0.01$), and also the 3-way interaction load x session x performance group ($F(3, 7) = 5.65$; $p < 0.01$). Post-hoc comparisons revealed that these interactions resulted from the performance by the low-performing group: This group showed some significant practice-related improvements in the second fMRI-session, i.e., they scored significantly higher in the 3-back tasks ($t(3, 7) = -4.39$; $p < 0.05$). In the 1-back task however, their performance significantly decreased ($t(3, 7) = 4.16$; $p < 0.05$). Conversely, the high-performing group did not show any significant performance differences between the two sessions on either load level (see Figure 10). The 3-way repeated measures ANOVA for RTs resulted in a significant main effect of load ($F(1.20, 7) = 25.92$; $p < 0.001$) and a main effect of session ($F(1, 7) = 7.97$; $p < 0.05$) with participants responding faster in the first session. There was no main effect for performance group, and no significant interactions were observed.
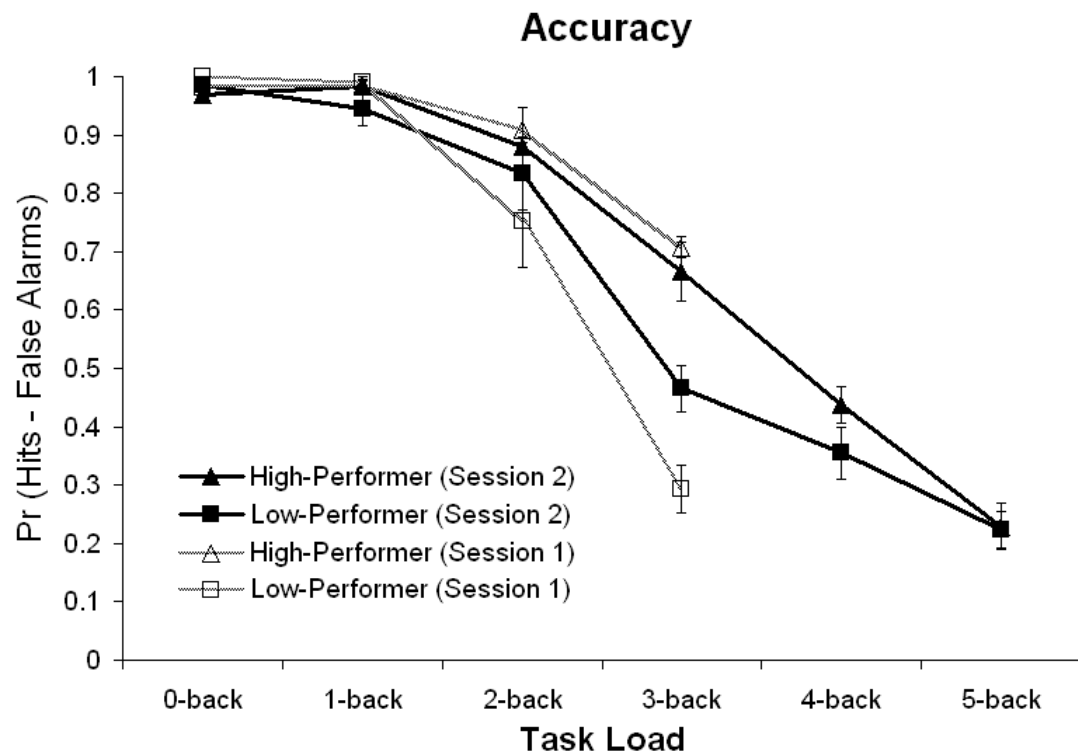
Figure 10. Means and SEM for the two performance groups in each session for each level of load. Since there were no 4- and 5-back tasks in session 1, only data for session 2 are plotted for these levels of load.

Imaging Data. On Table 9, the mean activation changes in relation to the baseline are shown for all defined VOIs, selectively for the high- and low-performing individuals and for each hemisphere.

Table 9. Activation changes in all defined volumes of interest (VOIs) for each level of load (compared to the baseline task, i.e., 0-back) and for each hemisphere, reported for both groups of participants in session 2.

| Brain Area | Parcellation Units (VOIs) | Abbr. | Left Hemisphere | | | | | | | | | | Right Hemisphere | | | | | | | | | |
| | | | Low Performer | | | | | High Performer | | | | | Low Performer | | | | | High Performer | | | | |
			1-back	2-back	3-back	4-back	5-back	1-back	2-back	3-back	4-back	5-back	1-back	2-back	3-back	4-back	5-back	1-back	2-back	3-back	4-back	5-back
Lateral cerebral surface																						
Frontal lobe	Central sulcus	ce	0.02	0.10*	0.14**	0.05	-0.12*	0.10	-0.01	-0.29***	-0.05	0.07	0.07	0.13**	0.10*	-0.03	-0.01	0.08*	-0.21***	-0.31***	-0.09*	0.03
	Superior frontal gyrus	F1	-0.01	0.12*	0.16**	0.26***	-0.02	0.00	-0.03	-0.15*	-0.11*	-0.25**	0.10	0.51***	0.58***	0.54***	0.39***	0.01	0.04	-0.06	-0.06	-0.08***
	Middle frontal gyrus (DLPFC)	F2	0.24***	0.60***	0.46***	0.47***	0.14*	0.19***	0.10	-0.01	0.10*	0.03	0.24***	0.44***	0.41***	0.39***	0.23***	-0.07*	-0.01	-0.06	-0.05	-0.16
	Frontal eye fields	FEF	0.01	0.23***	0.17**	0.36***	0.20**	0.12**	0.11	0.27***	0.14*	0.02	0.15**	0.26***	0.35***	0.21***	0.24***	-0.01	-0.03	0.13**	0.18**	-0.06
	Frontal pole	FP	-0.01	0.17	0.31***	0.12*	0.16**	-0.22**	-0.22*	-0.53***	-0.37***	-0.48***	-0.04	0.31*	0.55***	0.11	0.28**	0.06	-0.33**	-0.25**	-0.11	-0.10
	Inferior frontal sulcus	if	0.18**	0.54***	0.47***	0.44***	0.16*	0.17***	0.17**	0.09	0.10*	0.01	0.20**	0.54***	0.41***	0.46***	0.16**	0.05	-0.03	-0.06	-0.01	-0.16***
	Precentral sulcus[1]	prc	0.21***	0.47***	0.52***	0.49***	0.26***	0.18**	0.35***	0.07	0.28***	0.21*	0.31***	0.46***	0.34***	0.42***	0.36***	0.09*	0.11*	0.03	0.15***	0.10
	Superior frontal sulcus	sf	0.16**	0.40***	0.53***	0.51***	0.32***	0.14*	0.24**	0.10	0.19***	-0.05	0.17***	0.43***	0.41**	0.37***	0.30***	0.01	0.22***	0.15**	0.11*	0.00
Parietal lobe	Inferior intraparietal lobule[2]	INFP	0.22**	0.50***	0.69***	0.72***	0.40***	0.04	0.23	0.06	0.07**	0.11*	0.40***	0.69***	0.58***	0.51***	0.26***	0.08*	0.05	-0.07*	0.09**	0.16***
	Intraparietal sulcus	ip	0.34***	0.89***	0.84***	0.94***	0.50***	0.26***	0.52***	0.30***	0.33***	0.30***	0.34***	0.71***	0.62***	0.60***	0.34***	0.18***	0.29***	0.28***	0.25***	0.20***
	Postcentral sulcus	poc	0.11***	0.18*	0.32***	0.26**	0.01	0.22**	0.14**	-0.16**	0.01	0.12	0.17***	0.18***	0.11*	0.14**	0.18***	0.05	-0.14	-0.08	0.00	0.03
Temporal lobe	Inferior temporal sulcus	it	0.08	0.22***	0.19***	0.35***	0.15	0.05	-0.19*	-0.07	-0.03	-0.33***	0.19***	0.16***	0.12*	0.02	0.11*	0.02	-0.34***	-0.08	-0.16***	-0.26***
	Superior temporal sulcus	st	0.06	-0.02	0.10*	0.21**	0.11*	0.00	-0.15**	-0.12**	-0.08*	-0.32***	0.08	0.16***	0.00	0.19***	0.06	0.08**	-0.11**	-0.03	-0.08**	-0.14***
	Temporal pole	TP	0.01*	-0.10	-0.14	0.42***	0.09	0.25**	-0.75***	-0.47***	-0.54***	-0.55***	0.33**	0.22	-0.51*	0.08	0.01	0.25**	-0.15	-0.46***	-0.07	-0.17
Occipital lobe	Intracalcarine cortex	CALC	0.06	0.26***	0.22***	0.25***	0.15**	0.20***	0.07	-0.02	0.05	-0.12**	0.09*	0.28***	0.22***	0.28***	0.18***	0.15**	0.02	0.00	0.04	-0.13**
	Lateral occipital cortex, inferior division	OLi	0.11***	0.25***	0.31***	0.45***	0.05	0.03	-0.01	0.19**	-0.10***	-0.46***	0.17	0.15**	0.10	0.32***	0.08*	0.13**	0.00	0.03	0.02***	-0.11**
	Lateral occipital cortex; superior division	OLs	0.39**	1.20***	1.09***	1.30***	0.56***	0.19***	0.41***	0.35***	0.28	0.18***	0.29**	1.02***	0.93***	1.04***	0.47***	0.12**	0.14**	0.24***	0.18	0.18**
	Occipital pole	OP	0.15	0.12	0.31	0.80***	-0.11	-0.07	-0.03	0.20**	-0.14	-0.66***	0.22**	0.06	0.27	0.42***	0.03	0.07	0.08	0.25**	0.09	-0.39***
Medial cerebral surface																						
Frontal lobe	Cingulate gyrus; anterior division	CGa	-0.03	-0.05	0.07	0.25***	0.18**	0.03	0.02	-0.16***	-0.07	-0.29***	0.10*	0.15**	0.11*	0.26***	0.00	-0.03	-0.19***	0.00	-0.09*	-0.14**
	Cingulate Gyrus, posterior division	CGp	-0.01	0.00	0.12*	0.11*	0.14*	0.04	-0.21***	-0.21***	-0.20***	-0.20***	0.10	0.12**	0.18*	0.23***	0.07	-0.02	-0.37***	-0.38***	-0.27***	-0.36***
	Marginal ramus of cingulate sulcus	ma	0.07	0.07	0.12	0.30***	0.09	0.10*	0.10	-0.23***	0.04	0.03	0.01	0.17*	0.29***	0.23***	0.16*	0.05	0.11	-0.05	0.00	0.23***
	Paracingulate gyrus	PAC	-0.08*	0.15**	0.35***	0.26***	0.32***	0.05	0.22**	-0.10	0.11**	0.02	0.24***	0.67***	0.69***	0.42***	0.35***	0.11**	0.02	0.23***	0.21***	0.32***
	Supplementary motor cortex	SMC	0.06	0.21***	0.28***	0.03	0.05	-0.03	0.08	-0.33***	-0.16**	-0.22***	0.15	0.21**	0.23**	0.15*	0.08	0.13*	-0.18**	-0.21*	-0.16***	0.01

Table 9. (Continued).

Brain Area	Parcellation Units (VOIs)	Abbr.	Left Hemisphere										Right Hemisphere									
			Low Performer					High Performer					Low Performer					High Performer				
			1-back	2-back	3-back	4-back	5-back	1-back	2-back	3-back	4-back	5-back	1-back	2-back	3-back	4-back	5-back	1-back	2-back	3-back	4-back	5-back
Medial cerebral surface																						
Parietal lobe	Parietooccipital sulcus	po	0.18**	0.49***	0.42***	0.56***	0.43***	0.21***	0.34***	0.23***	0.19***	0.20***	0.21***	0.58***	0.57***	0.52***	0.31***	0.15***	0.14**	0.07	0.05	0.10*
	Subparietal sulcus	sp	0.02	0.07	0.18**	0.12*	0.10	0.03	-0.15*	-0.06	-0.16***	-0.26***	0.15*	0.20***	0.32***	0.21***	0.02	0.14**	-0.13	-0.13*	-0.05	-0.13**
Intrasylvian cerebral surface	Heschl sulcus	he	-0.06	0.04	-0.06	0.25***	-0.17*	-0.41**	-0.47***	-0.28**	-0.28***	-0.78***	0.06	0.13	-0.02	-0.03	-0.14*	-0.08	-0.47***	-0.18*	-0.16**	-0.33***
	Insular cortex	INS	0.02	0.11**	0.08*	0.22***	0.06	0.07*	-0.07**	-0.09*	0.02	-0.10*	0.01	0.16***	0.13***	0.19***	0.09*	0.01	-0.05	-0.17***	-0.02	-0.04
	Sylvian fissure (total extent)	sy	0.02	0.02	0.08	0.08	-0.14*	-0.04	-0.23***	-0.21***	-0.09*	-0.22***	-0.04	-0.04	-0.11*	0.00	-0.21***	-0.08	-0.40***	-0.30***	-0.33***	-0.48***
Subcortical structures	Caudate nucleus	cau	0.15**	0.41***	0.38***	0.61***	0.21**	0.09	0.08	-0.05	0.03	-0.34***	0.22**	0.28***	0.13	0.32	0.02	0.09	-0.11	0.18*	-0.08	-0.18*
	Putamen	pu	0.12*	0.11*	0.07	0.12**	0.03	0.04	0.07	0.06	-0.03	-0.14**	0.04	0.05	0.11*	0.27***	0.15**	0.01	-0.04	-0.13*	-0.06	-0.14**
	Thalamus	thal	0.05	0.17***	0.14***	0.20***	0.11**	0.08*	-0.04	-0.04	0.06	-0.03	0.11*	0.23***	0.10*	0.23***	0.00	0.07	-0.08	-0.11*	0.01	0.02
Cerebellum	Cerebellar hemisphere	CH	0.12*	0.21**	0.30***	0.43***	0.22**	0.02	-0.09*	0.18***	0.08*	-0.14**	0.05	0.21**	0.15*	0.44***	0.24**	-0.03	0.01	0.23***	0.13***	-0.14**
	Cerebellar vermis	CV	0.01	0.28***	0.20***	0.25***	0.28***	0.09	0.09	0.16***	0.14***	-0.09*	0.03	0.27***	0.21***	0.20***	0.16**	0.00	0.03	0.22***	0.04	-0.10*

Note: Values are given as group average of percent BOLD signal change (positive values = activation; negative values = deactivation) with statistical group analysis of fMRI data from VOIs. *p* values as two-tailed significance levels set at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[1] Premotor cortex adjacent to the precentral sulcus at the level of the "hand notch".

[2] Inferior parietal lobule comprising the intermediate sulcus of Jensen, the angular sulcus, and the anterior occipital sulcus.
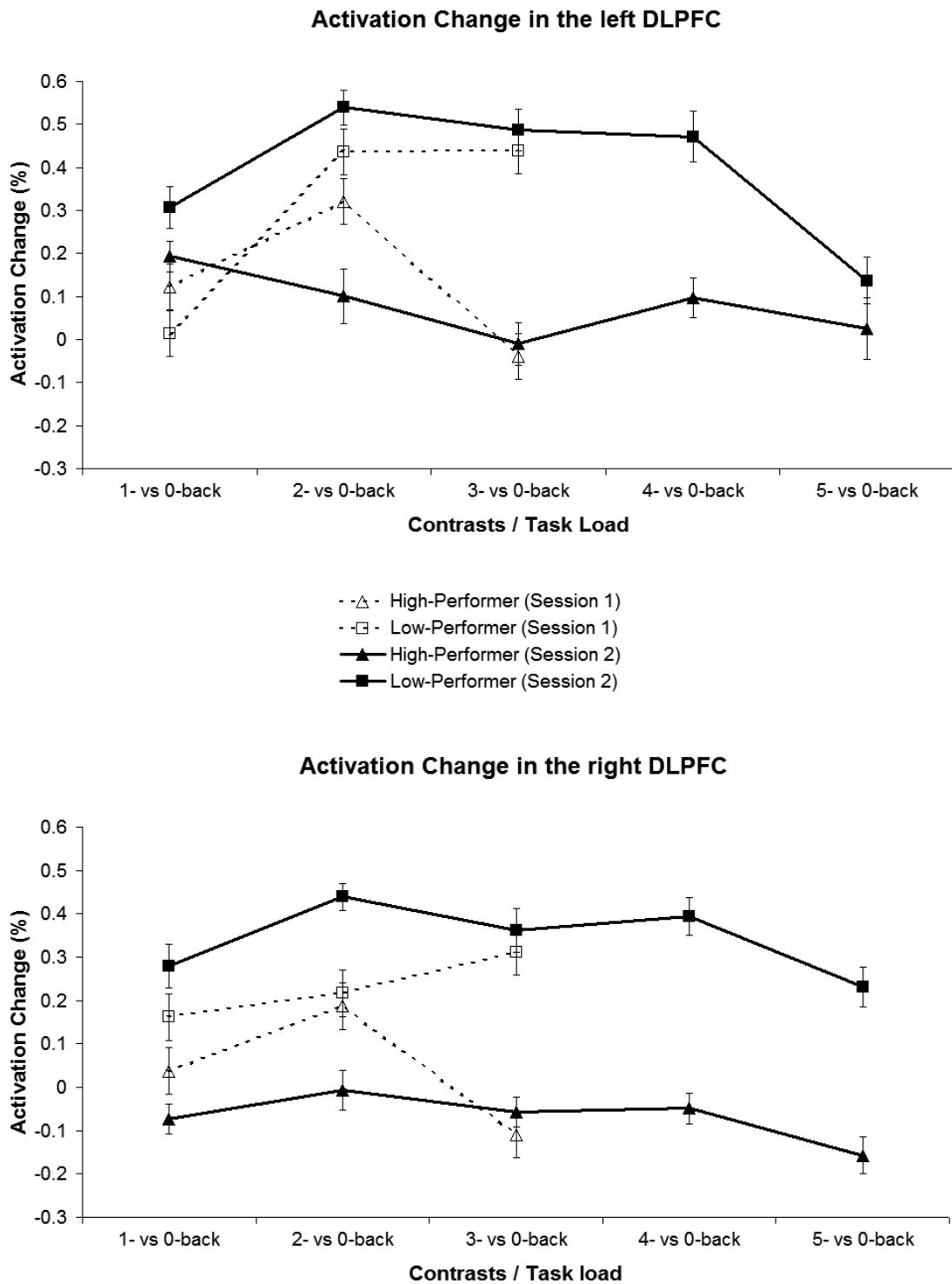
Concerning the DLPFC, the 3-way repeated measures ANOVA resulted in a significant main effect for load (1 to 5-back; $F(4,1375) = 11.24$; $p < 0.001$), a main effect for hemisphere (left vs. right; $F(1, 1375) = 18.39$; $p < 0.001$), a main effect for performance group (high vs. low performer; $F(1, 1375) = 243.87$; $p < 0.001$, as well as a significant interaction between load and performance group ($F(4, 1375) = 4.48$; $p = 0.001$), and hemisphere x performance group ($F(1, 1375) = 5.05$; $p < 0.05$). Neither the interaction load x hemisphere nor the 3-way interaction load x hemispheres x performance group was significant. Post-hoc tests indicated significantly higher activations in the low-performing group at each load condition, but with largest differences observed at intermediate levels, i.e., in the 2- to 4-back tasks (*1-back*: $t(211.01) = 5.25$; $p < 0.001$; *2-back*: $t(193.14) = 9.40$; $p < 0.001$; *3-back*: $t(222) = 9.89$; $p < 0.001$; *4-back*: $t(350) = 8.69$; $p < 0.001$; *5-back*: $t(340.35) = 4.53$; $p < 0.001$). Further, the low-performing group had higher activations in both hemispheres (*left*: $t(686) = 8.01$; $p < 0.001$; *right*: $t(686) = 14.98$; $p < 0.001$), but this difference was more expressed in the right DLPFC.

The 4-way repeated measures ANOVA with session as additional within-subject variable yielded a significant main effect of load (1 to 3-back; $F(2, 1343) = 17.98$; $p < 0.001$), a main effect for hemisphere (left vs. right; $F(1, 1343) = 21.42$; $p < 0.001$), a main effect for performance group (high vs. low performer; $F(1, 1343) = 176.52$; $p < 0.001$), but no main effect for session. The following 2-way interactions were significant: Load x performance group ($F(2, 1343) = 21.37$; $p < 0.001$), performance group x session ($F(1, 1343) = 23.11$; $p < 0.001$), and also the 3-way interactions load x hemisphere x performance group ($F(2, 1343) = 5.32$; $p < 0.01$), load x hemisphere x session ($F(2, 1343) = 3.27$; $p < 0.05$), load x performance group x session ($F(2, 1343) = 6.27$; $p < 0.01$). All other interactions were not significant. Post-hoc tests indicated that the difference between high- and low-performing groups observed in session 1 was even more pronounced in session 2 (*session 1*: $t(670) = 5.16$; $p < 0.001$; *session 2*: $t(670) = 14.04$; $p < 0.001$) with the low-performing participants showing substantially higher activations, which is also expressed in the 3-way interaction load x performance group x session, where the largest group-differences were observed at the 3-back tasks, but also, other than in session 1, in session 2 there were group-differences at every level of load (see 3-way ANOVA for group-differences above). The other 3-way interaction load x hemisphere x performance group revealed that the low-performing group showed higher activations as stated before, but also more expressed at higher levels of load, with the largest group-difference in the left

hemisphere in the 3-back task (significant differences between performance groups: *Left DLFPC*: 2-back: $t(222) = 5.21$; $p < 0.001$; 3-back: $t(222) = 8.87$; $p < 0.001$; *right DLPFC*: 1-back: $t(222) = 4.84$; $p < 0.001$; 2-back: $t(206.18) = 4.87$; $p < 0.001$; 3-back: $t(222) = 7.85$; $p < 0.001$). Generally, in session 2 there was more activation in the left hemisphere in the 1-back task than in the right hemisphere ($t(222) = -3.64$; $p < 0.001$), while at the other levels of load, no differences between hemispheres were observed, as was expressed with the interaction load x hemisphere x session.

The results of the ANOVAs for the DLPFC for both performance groups are plotted on Figure 11, and ANOVAs for the other VOIs are given on Table 10.

Activation Change in the left DLPFC

- - △ - High-Performer (Session 1)
- - ◻ - Low-Performer (Session 1)
- ▲ High-Performer (Session 2)
- ■ Low-Performer (Session 2)

Activation Change in the right DLPFC

Figure 11. Mean BOLD signal changes (in %) and SEM in the DLPFC, as difference between each load condition and the baseline-condition (0-back), plotted for the two performance groups and both hemispheres, with session 1 as plotted as reference. Since there were no 4- and 5-back tasks in session 1, only the data for session 2 are indicated for these levels of load.

Table 10. Main effects and interactions of the 3-way ANOVA shown for each defined VOI (alphabetically) in Session 2.

VOI	Load			Hemisphere			Performance group			Load x Hemisphere			Load x Performance group			Hemisphere x Performance group			Load x Hemisphere x Performance group		
	F	df	p	F	df	p	F	df	p	F	df	p	F	df	p	F	df	p	F	df	p
CALC	9.49	4	0.000	0.53	1	0.465	71.36	1	0.000	0.31	4	0.872	11.19	4	0.000	3.78	1	0.052	0.25	4	0.909
CGa	9.51	4	0.000	1.01	1	0.315	74.43	1	0.000	0.86	4	0.489	10.33	4	0.000	0.45	1	0.502	11.01	4	0.000
CGp	5.95	4	0.000	1.01	1	0.315	204.28	1	0.000	1.83	4	0.121	8.53	4	0.000	21.29	1	0.000	0.70	4	0.593
CH	13.56	4	0.000	0.01	1	0.930	102.65	1	0.000	0.47	4	0.756	7.96	4	0.000	1.60	1	0.206	0.48	4	0.748
CV	8.28	4	0.000	3.53	1	0.060	52.34	1	0.000	0.73	4	0.574	7.33	4	0.000	0.00	1	0.950	0.89	4	0.470
F1	3.38	4	0.009	40.53	1	0.000	76.51	1	0.000	1.92	4	0.104	6.35	4	0.000	15.28	1	0.000	0.47	4	0.757
F2	11.24	4	0.000	18.39	1	0.000	243.87	1	0.000	0.58	4	0.674	4.48	4	0.001	5.05	1	0.025	2.19	4	0.068
FEF	4.75	4	0.001	0.16	1	0.685	35.24	1	0.000	0.41	4	0.805	2.54	4	0.038	8.93	1	0.003	3.75	4	0.005
FP	1.70	4	0.148	14.78	1	0.000	98.26	1	0.000	1.86	4	0.115	7.01	4	0.000	3.28	1	0.070	1.31	4	0.264
INFP	4.02	4	0.003	0.52	1	0.471	186.15	1	0.000	1.79	4	0.128	7.36	4	0.000	0.50	1	0.481	4.80	4	0.001
INS	8.92	4	0.000	0.05	1	0.830	54.69	1	0.000	1.08	4	0.363	3.24	4	0.012	1.76	1	0.185	0.68	4	0.603
OLi	20.59	4	0.000	0.05	1	0.820	80.61	1	0.000	6.01	4	0.000	9.63	4	0.000	8.67	1	0.003	1.44	4	0.220
OLs	19.79	4	0.000	11.82	1	0.001	237.26	1	0.000	0.75	4	0.555	11.67	4	0.000	0.59	1	0.443	0.18	4	0.951
OP	28.24	4	0.000	3.04	1	0.082	40.92	1	0.000	1.61	4	0.168	11.44	4	0.000	4.60	1	0.032	2.19	4	0.068
PAC	6.90	4	0.000	55.69	1	0.000	40.08	1	0.000	2.19	4	0.068	3.91	4	0.004	7.90	1	0.005	10.28	4	0.000
SMC	3.75	4	0.005	2.81	1	0.094	81.46	1	0.000	2.73	4	0.028	5.90	4	0.000	0.01	1	0.910	2.12	4	0.076
TP	11.95	4	0.000	9.85	1	0.002	21.53	1	0.000	2.00	4	0.092	5.28	4	0.000	5.47	1	0.019	3.30	4	0.011
cau	11.66	4	0.000	4.55	1	0.033	64.31	1	0.000	1.52	4	0.193	7.07	4	0.000	7.66	1	0.006	2.66	4	0.031
ce	6.07	4	0.000	1.08	1	0.299	32.71	1	0.000	1.00	4	0.405	17.32	4	0.000	3.62	1	0.057	1.65	4	0.160
he	12.59	4	0.000	8.53	1	0.004	96.97	1	0.000	4.02	4	0.003	3.18	4	0.013	10.44	1	0.001	2.56	4	0.037
if	9.45	4	0.000	4.94	1	0.026	108.73	1	0.000	0.26	4	0.903	4.78	4	0.001	11.59	1	0.001	0.61	4	0.653
ip	18.47	4	0.000	32.07	1	0.000	108.97	1	0.000	1.19	4	0.312	7.64	4	0.000	2.33	1	0.127	1.59	4	0.175
it	5.96	4	0.000	3.99	1	0.046	114.35	1	0.000	4.19	4	0.002	6.92	4	0.000	0.00	1	0.951	1.58	4	0.178
ma	1.99	4	0.093	4.12	1	0.043	19.32	1	0.000	3.19	4	0.013	6.27	4	0.000	0.00	1	0.973	0.57	4	0.686
po	5.92	4	0.000	2.24	1	0.135	107.27	1	0.000	1.37	4	0.241	6.80	4	0.000	15.05	1	0.000	2.62	4	0.033
poc	2.00	4	0.092	3.16	1	0.076	27.70	1	0.000	0.91	4	0.459	3.71	4	0.005	2.52	1	0.113	5.35	4	0.000
prc	3.99	4	0.003	7.80	1	0.005	73.13	1	0.000	0.97	4	0.422	2.07	4	0.083	2.94	1	0.087	1.74	4	0.139
pu	3.44	4	0.008	1.57	1	0.210	31.05	1	0.000	3.03	4	0.017	4.23	4	0.002	4.00	1	0.046	0.78	4	0.537
sf	9.80	4	0.000	3.69	1	0.055	94.46	1	0.000	1.01	4	0.401	3.30	4	0.011	0.71	1	0.401	0.93	4	0.444
sp	4.95	4	0.001	15.69	1	0.000	62.07	1	0.000	1.01	4	0.400	3.72	4	0.005	2.24	1	0.135	3.55	4	0.007
st	10.99	4	0.000	4.75	1	0.029	84.52	1	0.000	0.83	4	0.503	7.93	4	0.000	4.10	1	0.043	2.30	4	0.057
sy	12.78	4	0.000	18.28	1	0.000	43.90	1	0.000	1.57	4	0.179	1.40	4	0.233	4.73	1	0.030	0.52	4	0.720
thal	5.99	4	0.000	0.95	1	0.330	44.09	1	0.000	0.27	4	0.900	4.84	4	0.001	0.09	1	0.767	1.89	4	0.110

*Note*: $N = 1376$ (number of measurements); see Table 9 for abbreviations.

Test-Retest Reliability of the BOLD-response (Pearson's Product Moment Correlations; *r*) in the various levels of load in the DLPFC yielded average to high values, with the highest correlations obtained in the 3-back task (see Table 11). These correlations were reliable, as indicated by the 95% confidence interval for the bootstrap test (3-back task; *left DLPFC*: $P(0.54 \leq 0.88 \leq 0.99)$; *right DLPFC*: $P(0.37 \leq 0.80 \leq 0.99)$).

Table 11. Test-Retest Reliability (*r*) for the activation changes (%) in the left and right DLPFC between the two sessions in the various levels of load.

		Session 2					
		Left Hemisphere			Right Hemisphere		
Session 1		**1-back**	**2-back**	**3-back**	**1-back**	**2-back**	**3-back**
Left Hemisphere	**1-back**	0.28					
	2-back		0.42				
	3-back			0.88**			
Right Hemisphere	**1-back**				0.48		
	2-back					0.20	
	3-back						0.80*

*Note. N = 8; * p < 0.05; ** p < 0.001*

Product Moment Correlations (Pearson's) were calculated to estimate the relationship between the BOLD-response and the performance measures in the DLPFC, yielding a substantial negative correlation in the 3-back dual task in both hemispheres (*right DLPFC*: $r = -0.55$; $p < 0.01$; *left DLPFC*: $r = -0.45$; $p < 0.05$), which is depicted in Figure 12. The 95% confidence intervals for the bootstrap test on the correlations show, that the measures are also highly reliable (*right DLPFC vs. P_r*: $P(-0.96 \leq -0.55 \leq -0.20)$; *left DLPFC vs. P_r*: $P(-0.99 \leq -0.45 \leq -0.03)$). In the other levels of load, the correlations were much smaller and all of them did not reach significance (*Left DLPFC*: 1-back: $r = -0.07$; 2-back: $r = 0.08$; *Right DLPFC*: 1-back: $r = -0.16$; 2-back: $r = 0.06$). For the 3-back task, also a substantial and reliable positive correlation between the left and the right DLPFC was observed ($r = 0.67$; $p = 0.001$; $P(0.32 \leq -0.67 \leq 1.00)$).
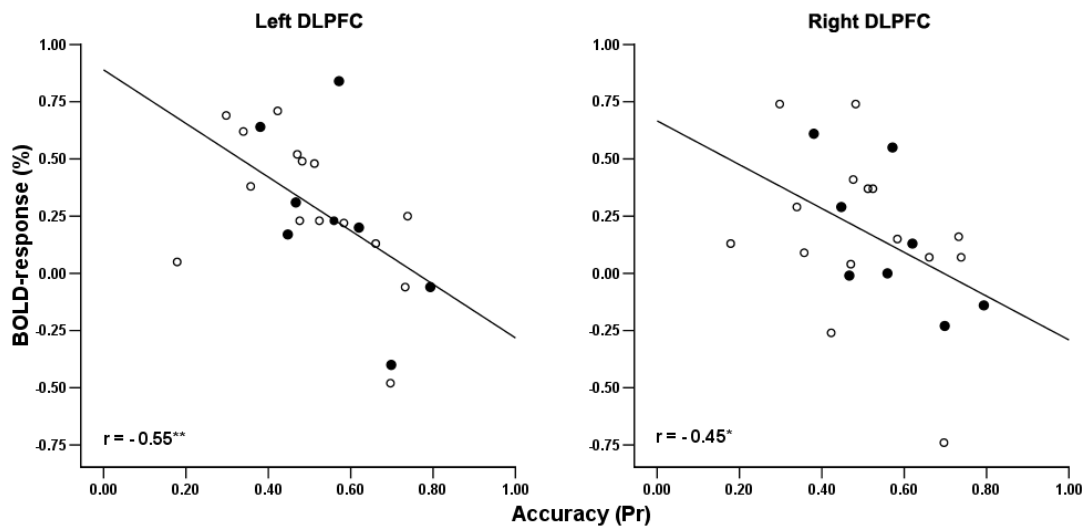
Figure 12. Scatterplots of accuracy (Pr) and BOLD-response in the DLPFC for each hemisphere, depicted for the 3-back task and both sessions (o = session 1; • = session 2); * $p < 0.05$; **$p < 0.01$; $N = 23$. As accuracy increased, the BOLD-response decreased as was expressed with the significant negative correlations. These correlations were all reliable at the 95% confidence interval, as was the positive correlation between the left and the right hemisphere ($r = 0.67$; $p = 0.001$).

4.4.4. Discussion

In this experiment, it has been shown that the forebrain of good performers indeed kept 'cool' in terms of brain activity even in conditions of excessive cognitive overload, although a decrease in performance was also observed in this group: The performance was no more different between the two groups in the 5-back task. Again, the data indicate that there seems to be no need for additional activation in prefrontal cortices to perform well in complex tasks, which might be explained in terms of a successful personality-related strategy, or the task-adequate and efficient functioning of neural circuits. The activation patterns in the DLPFC show that the brain of high-performing participants did not differentiate between high- and low-load conditions; rather, they seemed to equally distribute resources irrespective of task demands, apparently the most efficient strategy.

The reliability of the BOLD-signal and of the performance-measures (see also Chapter 3.2.1.1.), especially at high levels of load adds to the validity of the results and suggests that participants rely on the same strategies over different measurement-times, which is astonishing bearing in mind that the two sessions took place approximately 6 months apart, but which is nevertheless consistent with other findings in the literature, reporting large interindividual differences in the BOLD-response, which are however stable over time (Miller et al., 2002).

The negative correlations between the BOLD signal and the accuracy measures in the 3-back task provide further evidence that the DLPFC seems only to play a role if the participants are struggling with the task and may be interpreted as evidence against a *qualitative* difference between the high- and the low-performing group; rather, the better the performance, the less demands seem to be placed on prefrontal areas, which is expressed gradually over all participants. This finding is consistent with training studies, which usually report an activation decrease in task-related cortical brain regions after the task has undergone substantial practice. Such practice-related activation decreases have been demonstrated for example with motor learning (Van Der Graaf, De Jong, Maguire, Meiners, & Leenders, 2004), procedural learning (Kassubek, Schmidtke, Kimmig, Lucking, & Greenlee, 2001), but also in tasks relating to higher-cognitive functions, such as WM (Jansma, Ramsey, Slagter, & Kahn, 2001), or mathematical problem solving (Qin et al., 2004). Inconsistent to these studies, there are reports of an activation increase after WM training (Olesen,

Westerberg, & Klingberg, 2004), however, their participants performance did not consistently improve comparing pre and post sessions. It is argued, that the commonly observed activation decreases reflect a shift from controlled to automatic processing (Jansma et al., 2001), which does less rely on resources in task-related brain regions and thus reflects neuronal plasticity. Therefore, it could be assumed that some participants are faster in learning a task, such as the n-back task, relying on already established neuronal circuits, which is expressed in the lower BOLD signal in task-related brain areas, and in the observed interindividual differences. It seems therefore contradictory to these assumptions that there was no significant activation decrease in the low-performing participants from session 1 to session 2, despite their significant performance improvement in the 3-back task. This finding suggests that practice-related activation changes might require a longer time of training to be reflected on neural level and in the BOLD-response. It has to be noted, that this study was not designed as a training study: The two sessions took place 6 months apart without training in-between. Therefore, the high activation in the low-performing group might reflect the same executive strategies in both sessions, being nevertheless more efficient in session 2, which is expressed in the behavioural data. It can be hypothesized that before activation decreases can be observed, effective strategies have to be reliably established, which seem to take more time in some participants than two sessions in this task. The findings reported by Olesen et al. (2004) may similarly reflect that their participants are still in the process of strategy formation, which demands more neuronal resources reflected in increased cortical activity. It would be interesting to see, whether practice over a longer period of training would have reliable effects on the performance in the n-back task, and, related to the data presented here, whether the activations in low-performing participants would resemble those of high-performing participants after the training. To investigate some of the issues raised above, two training studies on behavioural basis will be presented in Chapter 6.

To summarize, the results of Experiment 2 concerning the interindividual differences in the BOLD-response in the DLPFC could be replicated, i.e. there was higher activation in the low-performing group than in the high-performing group, which reflects a most efficient recruitment of strategies and brain circuits in the high-performing individuals in tasks loading high on processing demands.

5. Capacity limitations and age

5.1. Age-related effects on processing capacity

There is a whole body of research providing evidence that certain cognitive functions decrease with age, while others are fairly well preserved. While abilities relying on knowledge and skills remain relatively stable until very old age, tasks related to 'fluid' processing (see Chapter 5) are much more vulnerable to the aging process (Park et al., 2002). Such observed cognitive declines are also reflected in age-related neuronal changes in the brain (Raz, 2005). However, the age-related neuronal and functional changes in the brain observed for example with fMRI are not well understood. While there are some brain areas in old adults where a reduction in activity is observed along with poorer performance, there are other findings where there are higher activities in some task-related areas compared to those of young adults (see Grady, McIntosh, & Craik (2005), or Della-Maggiore et al. (2002) for a review). These higher activations have been interpreted in the way that 'the older nervous system may functionally compensate for loss of function' (Della-Maggiore et al., 2002; p. 168). Despite these assumed compensation mechanisms, cognitive deficits clearly remain, and the nature of this behavioural performance decrease is not clear, as well as the age, when such decreases in performance are first pronounced, seeming variable and dependent on the type of task investigated (Park et al., 2002). Most studies are conducted with people at the age of 60 and above, but middle-aged adults (i.e., the age-segment of 50-60) are not well explored, despite accumulating evidence that changes in processing capacity are observed already much earlier (Lövdén & Lindenberger, 2005). The next experiment is now concerned with the question, how capacity-limits might change in advancing age and how effects of task load might be differentially expressed in two age-groups, first on behavioural basis, and second, by means of fMRI.

5.1.1. Working memory capacity in middle-aged adults as assessed with behavioural and functional data: Experiment 4a and 4b

The experiment consists of two parts: Part 1 (Experiment 4a) was conducted together with Corinne Schmid, who wrote her master thesis on the topic (Schmid, 2004). There healthy young and middle-aged participants were tested on behavioural basis regarding load-dependent processing by means of the n-back task at different levels of load as single and dual task. Part 2 (Experiment 4b) was an extension of Experiment 4a, investigating the neural correlates of capacity limits in a small healthy middle-aged sample by means of fMRI.

It was hypothesized that especially at high levels of load and in the dual-task conditions, group-differences would be observed, since in these conditions, executive functions being most vulnerable to aging-processes, are likely to be tapped.

The next question addressed the possible neural correlates of strategies, i.e. whether compensation mechanisms would be observable on functional basis. Relying on assumptions by Della-Maggiore et al. (2002), my hypothesis was that in case of similar performance with the young sample (i.e., probably with the low-performing group), the activation patterns in the middle-aged group would still be higher than those of the young participants in Experiment 2, i.e. reflecting the assumed compensation strategies.

Another hypothesis investigated concerned the HAROLD model (hemispheric asymmetry reduction in the old) introduced by Cabeza (2002): Taken the left lateralization observed in the young adults, the HAROLD model would predict compensatory mechanisms taking place in recruiting contralateral brain areas. Therefore, I expected bilateral activation in the middle-aged group, compared to the observed left-lateralized activation in the young sample, also provided that performance was comparable between-groups.

5.1.2. Method

Participants. Thirty middle-aged (9 men, 21 women; age range 50 to 64 years, mean age: 55.6) and thirty young participants (3 men, 27 women; age range 19 to 28 years, mean age: 21.78) took part in the behavioural study (Experiment 4a), whereas 7 middle-aged participants (3 women, 4 men; age range 51-63; mean age: 56.43) took part in the fMRI-study (Experiment 4b); however, one participant had to be excluded from the fMRI-data analyses because of a scanner-failure. The middle-aged participants volunteered for the experiment and were mainly recruited through the 'Altersuniversität' and the 'Volkshochschule' in Bern. The young sample consisted of undergraduate psychology students who participated for partial fulfilment of course requirements. All participants had normal or corrected-to-normal vision, good hearing, and no neurological disorders. Education level was the same in both groups, with a minimum of 12 years.

Material and Procedure. The same n-back task with the same material was used as described in Experiment 2.

Analysis.

Behavioural data: In Experiment 4a, 4-way repeated measures ANOVAs were calculated for the behavioural data, with task (single vs. dual), load (0 to 3-back) and modality (visuospatial vs. auditory) as within-subject factors and age-group as between-factor. In Experiment 4b, the modality-factor was not taken into account since these data were taken as basis to interpret the functional results, in which the different stimulus-materials processed in a dual-task cannot be selectively tackled.

Imaging data: For the imaging data, the DLPFC, defined as described in Experiment 1 will be the only VOI reported here, in order to keep it short and because the main hypotheses concerning differential neural effects in young and middle-aged participants had also been concentrating on this area (for a review, see Della-Maggiore et al., 2002). A 3-way repeated measures ANOVA was calculated, with load and hemisphere as within-factors and performance and age-group as between-factor.

5.1.3. Results

Behavioural experiment (4a)

All participants performed well in the dual-tasks, and performance was above chance level for all tasks in all conditions, even for the 3-back task under dual-task conditions ($t(58) = 0$; $p < 0.001$).

A 4-way repeated measures ANOVA on discrimination indices (P_r, i.e. hits minus false alarms; see Experiment 1) with task (single vs. dual tasks), modality (auditory-verbal vs. visuospatial-nonverbal) and load (0-back to 3-back) as within-subjects factors and age-group (young vs. middle-aged) as a between-subjects factor indicated significant main effects of task ($F(1,59) = 151.84$; $p < 0.001$), load ($F(3,59) = 9.39$; $p < 0.001$), modality ($F(1,59) = 19.60$; $p < 0.001$) and age-group ($F(1,59) = 5.40$; $p < 0.05$). There were also significant interactions of task x load ($F(2.33,59) = 44.72$; $p < 0.001$), modality x load ($F(2.20,59) = 39.67$; $p < 0.001$) and load x age-group ($F(3.29,59) = 3.21$; $p < 0.05$). The interaction task x age-group was also approaching significance ($F(1,59) = 3.63$; $p = 0.08$). Further, a 3-way interaction between task x modality x load ($F(2.10,59) = 7.46$; $p < 0.001$) was observed. As can be seen in Figure 13, accuracy rates were lower under dual-task conditions and decreased with increasing load, with all load levels differing significantly from each other ($p < 0.001$). Post-hoc comparisons revealed that this load-dependent decreases was larger in the dual tasks than in the single tasks and most expressed at the 3-back level (*1-back*: $Z = -4.80$; $p < 0.001$; *2-back*: $Z = -5.79$; $p < 0.001$; *3-back*: $Z = -6.53$; $p < 0.001$). Generally, the auditory tasks were more difficult than the visuospatial tasks, especially at higher levels of load, where the difference reached significance (*3-back*: $Z = -5.72$; $p < 0.001$). Load had a differential impact on the age-groups, with the middle-aged groups' performance being significantly lower than the performance of the young sample at the 2-back level ($\chi^2(1) = 7.52$; $p < 0.01$), with a trend for a difference also in the 3-back task ($\chi^2(1) = 3.07$; $p = 0.08$). The 3-way interaction task x modality x load indicated that the lower performance scores in the auditory tasks were more expressed at higher levels of load (2- and 3-back), and even more pronounced in the single tasks (single tasks: *1-back*: $Z = -2.71$; $p < 0.01$; *2-back*: $Z = -5.64$; $p < 0.001$; dual tasks: *0-back*: $Z = -3.35$; $p = 0.001$; *3-back*: $Z = -3.40$; $p < 0.001$).

The ANOVA for the RTs indicated significant main effects for task ($F(1,59)$ = 484.45; $p < 0.001$), load ($F(1.90,59)$ = 144.61; $p < 0.001$), and modality ($F(1,59)$ = 57.59; $p < 0.001$). No main effect for age-group was observed. There were two significant interactions, namely task x load ($F(2.52,59)$ = 38.72; $p < 0.001$), and modality x load ($F(2.04,59)$ = 5.17; $p < 0.01$). No other interactions were significant. Post-hoc tests indicated that the slower response times in the dual tasks were most pronounced at the 2-back level (*0-back*: $t(59)$ = -20.18; $p < 0.001$; *1-back*: $t(59)$ = -19.05; $p < 0.001$; *2-back*: $t(59)$ = -18.32; $p < 0.001$; *3-back*: $t(58)$ = -14.34; $p < 0.001$). The responses were also slower in the auditory tasks, with the largest differences at lowest and highest level of load (*0-back*: $t(59)$ = 9.21; $p < 0.001$; *1-back*: $t(59)$ = 3.68; $p = 0.001$; *2-back*: $t(59)$ = 4.46; $p < 0.001$; *3-back*: $t(58)$ = 4.97; $p < 0.001$), see Figure 14.
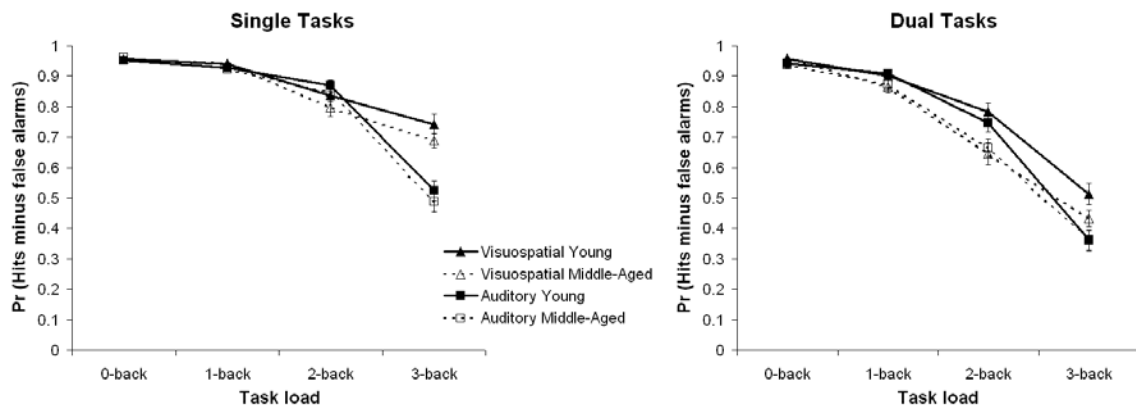

Figure 13. Accuracy measures (Pr). Means and SEM for the single and dual tasks, shown for each modality and age group.
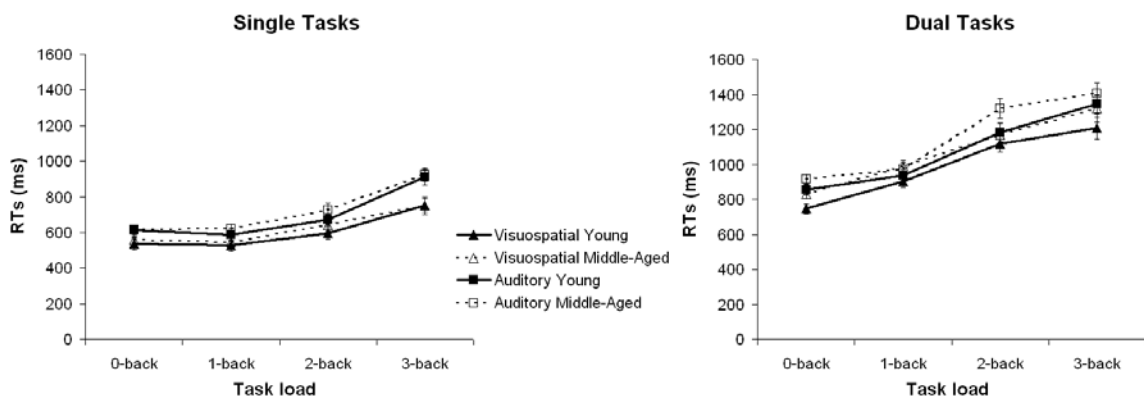

Figure 14. RTs. Means and SEM for the single and dual tasks, shown for each modality and age group.

fMRI experiment (4b)

  Behavioural data. The behavioural measures (P_r and RTs) of the middle-aged group were compared with those of the young sample in Experiment 3, which was divided into the performance groups as before (high, intermediate and low performer). The repeated measures ANOVA with load (0- to 3-back) as within-subject factor and performance and age group (high, intermediate, and low performer, as well as the middle-aged sample) as between-subject variable yielded a significant main effect of load (*accuracy*: $F(3, 21) = 172.74$; $p < 0.001$; *RTs*: $F(1.58, 21) = 30.71$; $p < 0.001$), and a main effect for performance and age-group (*accuracy*: $F(3, 21) = 9.12$; $p = 0.001$), as well as a significant interaction load x performance and age group (*accuracy*: $F(9, 21) = 8.16$; $p < 0.001$). Post-hoc tests showed that similar to Experiment 4a, the middle-aged group differed substantially from the young sample at higher levels of load, as expressed in significant differences between the middle-aged group and the high- and the intermediate performers on the 2- and 3-back levels (*2-back*: High performer vs. middle-aged: $t(8.04) = -4.52$; $p < 0.01$; intermediate performer vs. middle-aged: $t(7.89) = -3.88$; $p < 0.01$; *3-back*: High performer vs. middle-aged: $t(9) = -5.47$; $p < 0.001$; intermediate performer vs. middle-aged: $t(12) = -3.99$; $p < 0.01$). However, no difference was observed between the low-performing young participants and the middle-aged group at any level of load (see Figure 15).
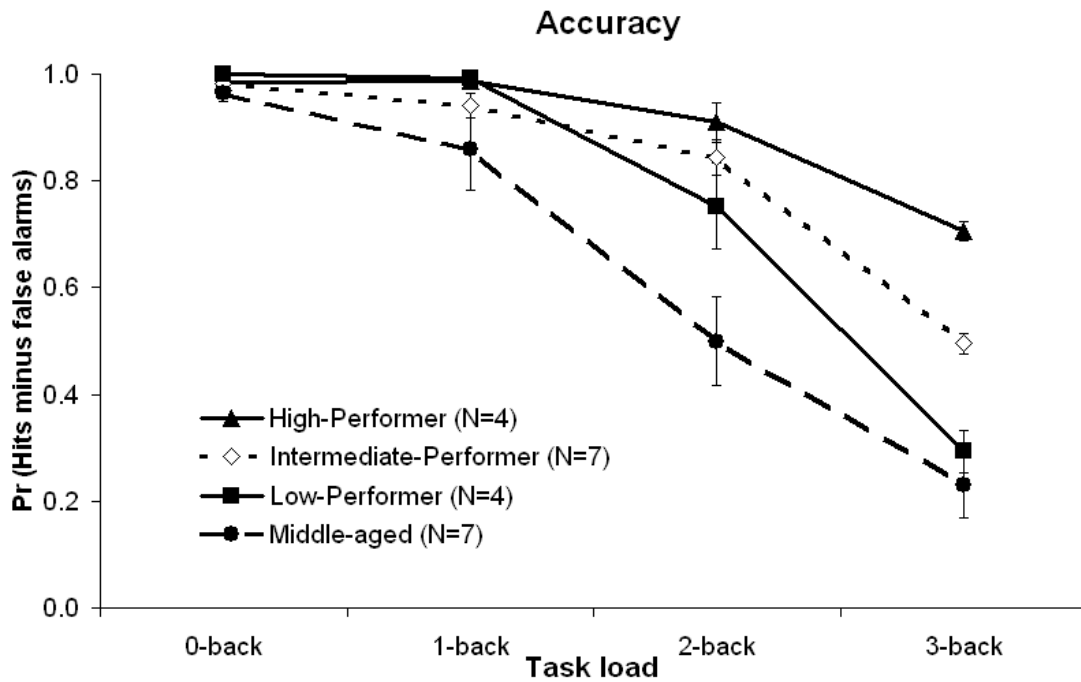
Figure 15. Mean performance levels (P_r) and SEM in the n-back dual task for each level of load, shown for the three performance groups of Experiment 2, including also the middle-aged sample from Experiment 4b.

Imaging data. The ANOVA with load (1 to 3-back) and hemisphere as within-subject factors, and performance and age-group (high, intermediate, and low performer, as well as the middle-aged) as between-subject variable indicated a main effect of load ($F(2, 1763) = 24.53$; $p < 0.001$), a main effect for hemisphere ($F(1, 1763) = 21.97$; $p < 0.001$), a main effect for performance and age group ($F(3, 1763) = 25.93$; $p < 0.001$). Significant interactions were observed between load x hemisphere ($F(2, 1763) = 5.07$; $p < 0.01$), load x performance and age group ($F(6, 1763) = 12.48$; $p < 0.001$). There was no interaction between performance and age group x hemisphere, and also, the 3-way interaction did not reach significance. Post-hoc tests revealed that the middle-aged group had higher activation changes in the 1-back task than all three young performance groups (middle-aged vs. high performer: $t(278) = -3.05$; $p < 0.01$; middle-aged vs. intermediate performer: $t(270.78) = -3.44$; $p = 0.001$; middle-aged vs. low performer: $t(278) = -3.09$; $p < 0.01$); and the same was true for the 3-back task (middle-aged vs. high performer: $t(278) = 9.55$; $p < 0.001$; middle-aged vs. intermediate performer: $t(221.83) = -3.44$; $p = 0.001$; middle-aged vs. low performer: $t(278) = -2.21$; $p < 0.05$). No group-differences were observed in the 2-back task (see Figure 16). For the group-differences in the young

sample see Experiment 2. Although the interaction hemisphere x performance- and age group was not significant, a selective pair-wise comparison was calculated for the middle-aged group, in order to test, whether bilateral activation would be observed exclusively in this group. However, also the middle-aged participants' left hemisphere showed significantly higher activation changes than their right hemisphere (t(252) = 5.51; $p < 0.001$).

Activation Change in the bilateral DLPFC
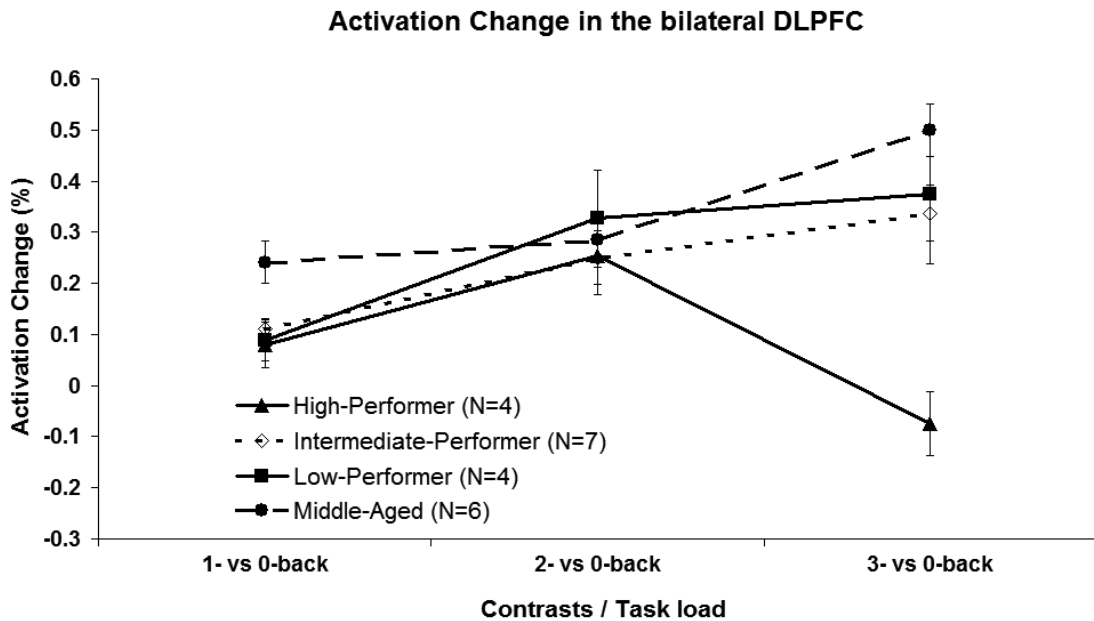
Figure 16. Mean BOLD signal changes (in %) and SEM in the bilateral DLPFC, as difference between each load condition and the baseline-condition (0-back), shown for the three performance groups in Experiment 2, but also including the middle-aged sample from Experiment 4b.

5.1.4. Discussion

The aim of this study was to investigate load dependent processing in a healthy, middle-aged and a young sample. As predicted, there were significant differences between the young and the middle-aged adults in both experiments, especially at higher levels of load, as being reflected with the significant interaction between load and performance and age groups in both behavioural experiments, and also, with the functional data.

In Experiment 4a, the observed age-related performance decline was most pronounced in the 2-back task and less for the most difficult level, i.e., for the 3-back task. This finding might account for a floor level effect on the 3-back tasks: Even though all participants performed above chance level, performance of the 3-back task might not differentiate as well as the 2-back task between the two age-groups. Conversely in Experiment 4b, the middle-aged group differed significantly from the young sample in the 3-back task, but not in the 2-back task. The difference in the 3-back task however, resulted from the difference between the middle-aged group and the high- and intermediate-performing young individuals, suggesting therefore that the middle-aged sample indeed performed as the low-performing participants as predicted and thus fulfilling the requirement in testing the hypothesis of a compensatory activation change in the PFC. The results showed that there were indeed significant differences between the young and the middle-aged sample, pronounced at lowest level of load (1-back) but also in the most difficult 3-back task: The middle-aged group showed markedly higher BOLD-responses in the DLPFC than all young performance-groups. This finding therefore supports the assumption of compensatory mechanisms, being pronounced with higher activation changes and is further validated in the equal performance level of the middle-aged and the low-performing participants, making it clear that the higher activation is not solely attributable to the worst performance in the middle-aged group. The same applies to the 1-back condition, where performance is the same across all groups, although the middle-aged group shows much higher activation changes in this condition. It could be assumed however, that age-related vascular changes contribute to this higher BOLD-signal (see Gazzaley & D'Esposito, 2005 for a review), though the comparable activation change between all groups in the 2-back task speaks against this interpretation.

The HAROLD model could not be supported with our data; rather, a left-lateralized activation in the DLPFC was observed also in the middle-aged sample. It

seems therefore, that compared to other memory tasks, and the conceptualization of a reduction in hemispheric asymmetry does not seem to apply to the used n-back task. It seems therefore that the assumed compensatory mechanisms do not result from recruiting contralateral areas, but from the recruitment of additional, but already activated neuronal circuits, therefore reflecting a quantitative rather than a qualitative age-related change in brain function.

There were no group-differences in terms of RTs in both experiments, providing further evidence that despite load-dependent increases, RTs do not reflect interindividual differences in this task. It is surprising however, that no differences between the young and the middle-aged sample was observed at the tasks at lower levels of load, i.e. in the 0- and 1-back task. According to the processing speed hypothesis, such simple perceptual speed tasks are supposed to be the best predictors of age-related differences in cognition (Salthouse, 1996; Salthouse & Miles, 2002). Therefore, the observation of a specific age-related performance decline without significant slowing-down in middle-aged participants suggests that there are other factors accounting for age-related decline in cognition.

Taken together, the results of Experiment 4a and 4b provide interesting insights in the domain of age-related changes in load-dependent processing in the middle-aged. Our results are in line with findings in the literature, suggesting a performance related decline in executive tasks over the life-course, already present at the age of 50 (e.g. Fisk & Warr, 1996; Meguro et al., 2000; Park et al., 2002; Salthouse et al., 2003). The functional data speak for quantitative compensatory mechanisms on neuronal level, being in line with the suggestions made by Della-Maggiore et al. (2002). Still, several limitations in the present study should be noted primarily related to the imaging data: Since I only investigated the DLPFC for this purpose it cannot be excluded that there might be other brain regions, where the HAROLD model might apply. Another limit is the small sample size, which performed quite homogeneously, i.e. similar to the low-performing group. It would be interesting to see, whether there are similar high activation patterns in middle-aged adults, which show a better performance than the low-performing group. Nevertheless, the results of these two experiments do certainly add to a better understanding of age-related changes in load-dependent processing and related brain function.

5.2. Summary and conclusion on capacity, neurophysiology, and age

What can be concluded from the last three experiments? I will summarize some of the most important findings:

There are large interindividual differences in load-dependent processing, especially in dual tasks, which are observed both on behavioural and on neuronal basis. These differences most likely result from the use of different strategies and the more efficient distribution of resources most pronounced in high-performing participants. This optimizing strategy is expressed in a negative correlation between performance and the BOLD-response, i.e. in a lower BOLD-signal in many cortical areas, but most consistently in the DLPFC. It is not clear though on what basis participants choose these strategies, whether they can be taught and made more efficient with training, and/or whether they are related to intelligence, a question which will be addressed in the next chapters.

In advancing age, there seem to be a declining cognitive function as assessed with the n-back task, probably related to executive processing, which is expressed in lower accuracy measures in the middle-aged sample at higher levels of load. The stronger BOLD-signal in this group might reflect compensatory mechanisms on neuronal level in order to fulfil the required task demands.

6. Capacity and intelligence

6.1. General issues

The relationship between mental test performance (i.e., intelligence) and information-processing abilities, that is, the cognitive correlates of ability measures has been investigated and debated for more than a century (for a review, see Sternberg (1977); or more recently Heitz et al. (2005)). Already Binet (1903) and Spearman (1927) discussed the idea that attention might have a considerable impact on intelligence, also with converging evidence from recent factor-analytical studies, 'that attention is one of the major determinants of fluid intelligence – the ability to reason abstractly in novel environments and situations.' (Heitz et al., 2005; p. 62). Similarly, Cowan et al. (2005) most recently stated, that the common variance between various measures and individual differences in attentional capacity do strongly relate to measures of intellectual aptitudes.

In this context, it is important to clarify what is meant with intelligence before it is talked about its relationship to other cognitive measures. In the psychometric study of intelligence, there are several approaches, which differ on important domains: The first view stems from Spearman (1904; 1927) with his two-factor theory of intelligence: There, the idea of 'positive manifold' is proposed, meaning that one single factor (i.e., *Spearman's g*, the general intellectual ability; i.e., a general factor that governs performance on all cognitive tasks) determines individual differences. Another factor (*s*) is specific for the tests being used, and therefore explaining the rest of the variance. A contrasting view is taken by theories, mainly represented by Thurstone (1938), assuming a wide range of independent 'primary abilities', i.e., factors, which are not seen as contributing to a general factor such as *g*. The third view, put forward by Cattell (1943; 1963) can be regarded as compromising the two approaches above, and proposes a hierarchical model, dividing *g* in two sub-components, i.e., general-crystallized (Gc) and general-fluid (Gf) intelligence. In this model, Gc represents the accumulated knowledge and abilities over time and can be measured for example with tests of factual knowledge, or vocabulary. Gf on the other hand can be described as the ability to deal with new situations, to solve problems, and for abstract reasoning, being measured with tasks usually described as 'higher-order cognitive tasks', such as matrices tests (e.g. Raven's Progressive Matrices), or

tests for executive functions (e.g. the tower of Hanoi). It is worthwhile to note that these three historic, but nevertheless still widely acknowledged theories of intelligence are derived factor-analytically, that is, in a data-driven, bottom-up way, which is in sharp contrast to the way in which for example WM research is tackled, i.e., primarily top-down and theory-driven. Of course, there are many other approaches to understand intelligence, but the differential discussion of all these would go beyond the scope of this thesis and I will therefore focus on the topic of the relationship between general intelligence in the sense of Spearman's *g* and WMC. However, as stated earlier, there is much controversy about the relationship of *g* and WMC, with the debate still lively going on in the literature, for example between the groups around Ackerman, Kane and Engle, and Oberauer (Ackerman, Beier, & Boyle, 2005; Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Suss, 2005). While some researchers provide evidence in favour of a separation of the two concepts (Ackerman et al., 2005), although willing to admit an association to some extent, others support the view of a strong correlation between the two (e.g. Kane et al., 2005; Kyllonen & Christal, 1990), which is also underpinned by the evidence that they seem to share a common neural substrate, namely the prefrontal cortex (e.g. Gray et al., 2003; for a review, see Kane & Engle, 2002). Usually, tasks representing reasoning and problem solving-abilities seem to be highly correlated with WMC, which can be interpreted that solving those tasks depends on being able to construct complex structural representations, which is in turn dependent on WMC (Halford et al., 1998; Johnson-Laird, 1983; Oberauer, 2005; Wilhelm, 2005).

Depending on whether accuracy measures or reaction times are taken into account, and also, whether easy or difficult, i.e., more resource-demanding tasks are administered in order to assess WMC, a differential impact on psychometric *g* can be observed. While some researchers argue, that RTs in very simple, non-intellectual tasks can serve as predictor of non-speeded complex tasks of general knowledge and reasoning abilities (e.g. Jensen, 1993; Vernon, 1983; Vernon & Jensen, 1984), others go as far as stating the speed of information processing as equivalent to *g* (Salthouse, 1996; Salthouse et al., 2003). In the context of WMC, Jensen (1993) argues that there is much evidence that RT tasks are more related to *g*, if capacity is strained to a certain extent. However, if capacity limits are reached or exceeded, this correlation will tend to decrease. Hence, again, an inverse U-shaped curve is proposed in regard

to RTs and their relationship to *g*: If the amount of information processing is very small, that is, if response times in simple tasks are measured, the correlation between RTs and *g* would be in the range of $r = - 0.10$. But with increasing processing complexity by increasing choice and discrimination, this correlation can be enhanced, and again, decreases if too many errors are made and hence, if capacity limits are reached. Therefore, the best measures of an individual's level of *g* are described as tasks *at* capacity limitations, but not *beyond* (Jensen, 1993). Experimental evidence to this complexity hypothesis comes from work with a dual task paradigm: Consistently greater correlations between RTs and *g* are observed in dual tasks compared to single-task conditions (Jensen, 1993), which can also be explained with the use of strategies, being more pronounced in single tasks, and therefore contributing to source of error variance (Oberauer, 2005). Further evidence that dual tasks load higher on *g* comes from earlier studies by Fogarty and Stankov (1982) and Spilsbury (1992), and is also acknowledged by Kane, Engle, and colleagues (e.g. Heitz et al., 2005; Kane, 2005; Kane & Engle, 2002), stating that dual-task situations place much more demands on executive control, i.e., 'executive attention' than single tasks, and therefore contributing much more to *g*.

However, working solely with RTs as measure of performance has some important drawbacks: Response latencies alone do not clarify the processes underlying the correlations between WMC and *g* – usually the discussions end up with the question, whether speed is a valid predictor of *g*, and therefore raise concerns about the underlying mechanisms of individual differences in speed (What are they? Nerve conduction velocity? Strategic processing mechanisms? Efficiency in retrieval? Other mechanisms?), as discussed for example in Heitz et al. (2005) or Conway (2005). Stankov and Crawford (1993) compared both, measures of accuracy and RT regarding their differential impact on *g* with a task of various levels of complexity. They found comparable correlations between the two behavioural measures and *g*, increasing linearly with the increase of complexity, with the strongest correlation at highest levels of complexity (accuracy: $r = 0.46$; RT: $r = -0.40$). In the view of Hockey and Geffen (2004) 'this suggests that with increasing levels of task complexity, participants' start using differential cognitive strategies, and that an individual's ability to store and manipulate increased amounts of information in working memory partially determines the efficacy of a particular strategy.' (p. 593).

These different strategies are probably not easily reflected in RTs only and favour the application of both measures, also to investigate their differential impact on *g*, especially at high levels of complexity.

Much body of research as to the relationship between WMC and *g* comes from quasi-experimental studies by the Kane, Engle and colleagues (e.g. Kane & Engle, 2000; Rosen & Engle, 1998), or Conway (e.g. Conway, Cowan, & Bunting, 2001), focusing mainly on WM span and divided attention tasks. In their work they investigated high and low WM span participants in their regard to their susceptibility to interference as tested with divided-attention tasks. Indeed, it could be shown that low-span participants are reliably more vulnerable to interference than high-span participants. Additionally, it was observed that high-span individuals 'simulate' low-span behaviour in divided-attention conditions, which is explained with the association between high WMC and executive attention capability, being disrupted in dual-task situations (Kane & Engle, 2000). In sum, '(…) tasks that require controlled attention to maintain information, inhibit distracter information, or block prepotent response tendencies seem to dissociate individuals scoring low or high on dual tasks (…).' (Heitz et al., 2005; p. 67).

6.1.1. Are capacity-measures and intelligence related constructs?

Experiments 5 and 1a

In the following experiments, I was interested whether the proposed relationship between WMC and fluid intelligence could also be confirmed with the n-back task as an indicator of WMC, and also, the differential impact of WMC on fluid, or more crystalline components of intelligence was explored. Further, an extreme group approach was adopted in order to investigate the impact of high and low performing individuals on intelligence. The data of two experiments were evaluated: In Experiment 6, the n-back task was used at 3 levels of difficulty as in Experiment 1 (1 to 3-back; single and dual tasks), and further, the Raven advanced progressive matrices (APM; Raven, 1990) were administered as an index of fluid intelligence. In Experiment 1a, in addition to the n-back task (single and dual task, 1 to 3-back), a vocabulary test, the 'Wortschatz-Test' (WST; Schmidt & Metzler, 1992) was used as an index of verbal, i.e., crystalline intelligence. In this experiment, the data reported here are supplementary to Experiment 1[8] and it will be therefore referred to as Experiment 1a.

Based on the findings from the literature as described in Chapter 5.1., the following hypotheses were investigated:

Correlating the n-back task with the APM, I expected a substantial correlation between the two measures, and further, the correlation between the n-back task and the intelligence measure should increase linearly in terms of accuracy (Hockey & Geffen, 2004; Stankov & Crawford, 1993). The correlations should be more pronounced in the dual-task conditions, because of the larger variability observed in the dual tasks, which would represent larger individual differences (see Experiments 1, 2 and 4), and because dual tasks seem to be better predictors of g (Fogarty & Stankov, 1982; Spilsbury, 1992). Also, since g is assumed to reflect a domain-free process (e.g. Kane & Engle, 2002), no differences in the correlations between the visuospatial and the auditory versions of the n-back task and the APM are predicted.

---

[8] The WST was administered later in the testing session as additional information, therefore, WST data are missing for the first 21 participants from the original sample ($n = 132$), resulting in an n of 101.

Applying Jensen's complexity hypothesis (Jensen, 1993) to the n-back task, it was predicted that the greatest correlation between the n-back task and g as assessed with the APM would be observed at intermediate levels of the n-back in terms of RTs.

For crystalline intelligence however, it was predicted that if any correlation would be observed between the n-back task and the crystalline intelligence measure, i.e., the WST, it would be pronounced at higher levels of load (2- and 3-back), and moreover, because it relies primarily on the verbal domain, more prominent correlations are expected in the auditory-verbal versions of the n-back task.

Further, the extreme group methodology as mentioned above was used again, as described in Experiment 3, i.e., groups of participants were differentiated based on their performance in the dual 3-back task (see method section below). Besides the dissociation in the n-back task, I expected that high-performing participants would also score higher in the APM than low-performing participants, taken into account that WMC would be the basis of higher intellectual abilities like reasoning as measured with the APM. On the other hand, similar to the approach adopted by Stankov and Crawford (1993) and also Hockey and Geffen (2004), I examined individuals scoring high and low in the APM in terms of differences in the various levels of the n-back task not necessarily consisting of the same participants as in the reverse relationship. I predicted that individuals with higher scores on the APM would also perform better in the n-back task, especially at high levels of load. Both of these relationships are less expected in regard to the WST, since there is much less evidence for a relationship between WMC and measures loading on crystalline intelligence (see for example Hockey & Geffen, 2004).

6.1.2. Method

Participants. 50 participants took part in Experiment 5 (26 women and 24 men; mean age: 20.44; *SD*: 3.56). Further, in Experiment 1a, the data of 101 participants (65 women, 36 men; mean age: 29.50; *SD*: 11.21), which were not the same as in Experiment 5, were included in the analyses. In both studies, most of the participants were recruited by undergraduate students in order to fulfil course credit and consisted of undergraduate students, college students and others.

Material and Procedure. N-back task: For Experiment 1a, the n-back task was used as described in Experiment 1, i.e. with three levels of load (1- to 3-back, single and dual task). In Experiment 5, also the same n-back task was used, but with four levels of load, i.e. with an additional 0-back task. Further, the APM (Raven, 1990) were used in Experiment 5, and in Experiment 1a, the WST (Schmidt & Metzler, 1992) was administered (see below).

APM (Raven, 1990): This test was developed as a measure of fluid intelligence. It consists of 36 visual pattern matching and analogy problems with increasing difficulty; that is, the correct pattern out of 8 possibilities has to be selected and matched to 8 presented patterns. According to Lezak, Howieson, and Loring (2004), in Ravens' matrices tests participants are required to 'conceptualize spatial, design, and numerical relationships ranging from the very obvious and concrete to the very complex and abstract' (p. 579). It is thought to be relatively language-free and to measure the individual's 'culture-free' ability to reason, being very well predictive for *g* (e.g. Spearman (1938); as cited in Ackerman et al., 2005) and the most generally accepted measure of fluid intelligence (Cowan et al., 2005). The correlation between the reasoning tasks such as the APM and WMC usually lies between $r = 0.54$ and $r = 1.00$ as reviewed in Kane et al. (2005), thus providing further evidence for the strong relationship between *g* and WMC (however, see Ackerman et al., 2005). The APM used in our experiment were developed especially in order to test participants with above average intellectual abilities. Therefore, regarding my group of participants, the overall level of difficulty was expected to be ideal, i.e., neither floor nor ceiling effects are expected.

WST (Schmidt & Metzler, 1992): In this task, participants have to select a word out of 5 pseudowords (example: Renek – Skerk – Erenk – Kern – Nerk – Lersk; correct answer: Kern), and the number of correctly identified words serves as measure

of verbal intelligence. The task consists of 42 trials with increasing difficulty. It was developed as indicator of crystalline intelligence, and is indeed strongly associated with scholar aptitude ($r = 0.60$) but not with age ($r = 0.08$; Schmidt & Metzler, 1992).

Data Analysis. Product Moment Correlations between the scores in the various levels of the n-back task and the intelligence measures were calculated. Also, extreme groups were created as in Experiment 3: High- and low-performing participants were selected based on their performance in the 3-back dual task, i.e., individuals performing in the lowest and highest quartiles were chosen as representing high and low performers. These two groups were compared in regard to their performance in the intelligence measures, as well as in the other n-back conditions. Additionally, high- and low-performing participants in the two intelligence-measures (similarly selected from the lowest and highest quartile) were compared in regard to their performance in the n-back task. For these comparisons, nonparametric tests where applied, where Kolmogorov-Smirnoff tests indicated non-normal distributions of the data.

6.1.3. Results

The descriptive data for the performance in the two intelligence measures as well as for the 3-back dual tasks for both experiments are shown in Table 12, and the correlations between the performance in the various levels of load in the n-back tasks and the two intelligence measures are indicated in Table 13.

Table 12. Means, standard deviations and percentile scores for the Raven Advanced Progressive Matrices (APM) and the verbal intelligence test (Wortschatztest; WST), as well as for the 3-back dual task (Experiment 5 and 1a).

				Percentiles		
		Mean	*(SD)*	25%	50%	75%
Experiment 5	APM	27.90	*(4.03)*	25.00	28.00	31.00
	3-back dual task					
	P_r	0.29	*(0.17)*	0.17	0.28	0.41
	RT	1245	*(417)*	940	1214	1577
Experiment 1a	WST	34.20	*(3.15)*	33.00	35.00	36.00
	3-back dual task					
	P_r	0.37	*(0.13)*	0.29	0.38	0.46
	RT	1339	*(396)*	1066	1290	1594

Note. The 3-back dual represents the mean between the visuospatial and the auditory task. Experiment 5: $N = 50$; Experiment 1a: $N = 101$. RTs are represented as the mean of median ms (hits only). *SD*: Standard deviation. P_r: Accuracy, i.e. hits minus false alarms.

Table 13. Pearson's Product Moment Correlations between the intelligence test scores and the various levels of the n-back task as assessed with accuracy and RTs in Experiment 5 and 1a.

Complexity level	APM (fluid intelligence; N=50)		WST (crystalline intelligence; N=101)	
Single Tasks	*Accuracy*	*RTs*	*Accuracy*	*RTs*
0-back auditory	-.14	-.08		
1-back auditory	-.13	-.23	.11	.06
2-back auditory	.17	-.15	.14	.00
3-back auditory	.29*	-.10	.06	.10
0-back visuospatial	-.13	-.09		
1-back visuospatial	-.02	-.16	.11	.05
2-back visuospatial	.26	-.10	.26*	.10
3-back visuospatial	.32*	-.19	.11	.11
Dual Tasks				
0-back auditory	-.15	-.08		
1-back auditory	-.21	-.04	.30**	.07
2-back auditory	-.05	-.01	.39**	-.08
3-back auditory	.12	-.02	.17	.16
0-back visuospatial	-.17	-.18		
1-back visuospatial	-.20	.05	.31**	.16
2-back visuospatial	.11	.05	.16	.16
3-back visuospatial	.60**	-.07	.15	.23*

Note. * $p < 0.05$; ** $p < 0.01$ (2-tailed). In the experimental group performing the WST, no 0-back task was applied, therefore, no data are available in regard to this task.

Experiment 5. The correlations between the APM and the n-back tasks yielded low to average correlations, with the visual 3-back dual task as highest correlation ($r = 0.60$; $p < 0.01$) obtained with accuracy measures as indicated in Table 13. With RTs, no significant correlations were observed between any level of the n-back task and the APM-scores.

As in Experiment 2, two groups of participants were selected, based on their performance levels in the 3-back dual task (lowest and highest quartile; see Table 12). The extreme group sample consisted of 11 low-performing participants ($P_r < 0.17$; $M = 0.08$; $SD = 0.08$) and 13 high-performing participants ($P_r > 0.41$; $M = 0.50$; $SD = 0.10$). As predicted, the APM scores differed significantly between the two samples: The mean APM score in the low-performing participants was 26.25 ($SD = 3.98$), whereas it was 31.31 ($SD = 3.17$) in the high-performing participants ($Z = -2.94$; $p = 0.001$; one-tailed). The two groups also differed on other levels of the n-back task in terms of accuracy, as illustrated in Figure 17 (*single tasks*: Auditory 1-back ($Z = -1.94$; $p < 0.01$); auditory 3-back ($Z = -2.38$; $p < 0.01$); visuospatial 3-back ($Z = -2.52$;

$p < 0.01$); *dual tasks*: Auditory 3-back ($Z = -3.74$; $p < 0.001$); visuospatial 3-back ($Z = -4.15$; $p < 0.001$); all one-tailed). In RTs, significant group differences were observed in some single tasks: In the visuospatial 1-back task ($Z = -1.77$; $p < 0.05$), in the visuospatial 2-back task ($Z = -2.41$; $p < 0.01$), and in the auditory 2-back task ($Z = -1.74$; $p < 0.05$; all one-tailed), in all of which the low-performing group responded faster.

If extreme groups were selected on the basis of APM scores, 13 participants were classified as low-APM (APM < 25; $M = 22.69$; $SD = 2.10$) and 17 as high-APM (APM > 31; $M = 32.29$; $SD = 1.45$). As can be seen on Figure 18, several differences are observed concerning the performance in the n-back tasks, which were throughout pronounced in the visuospatial tasks (*single tasks*: Visuospatial 1-back ($Z = -2.81$; $p < 0.05$), visuospatial 2-back ($Z = -1.88$; $p < 0.05$); visuospatial 3-back ($Z = -2.07$; $p < 0.05$); *dual tasks*: Visuospatial 3-back ($Z = -3.393$; $p < 0.001$); all one-tailed). No differences in terms of RTs were observed between these groups.
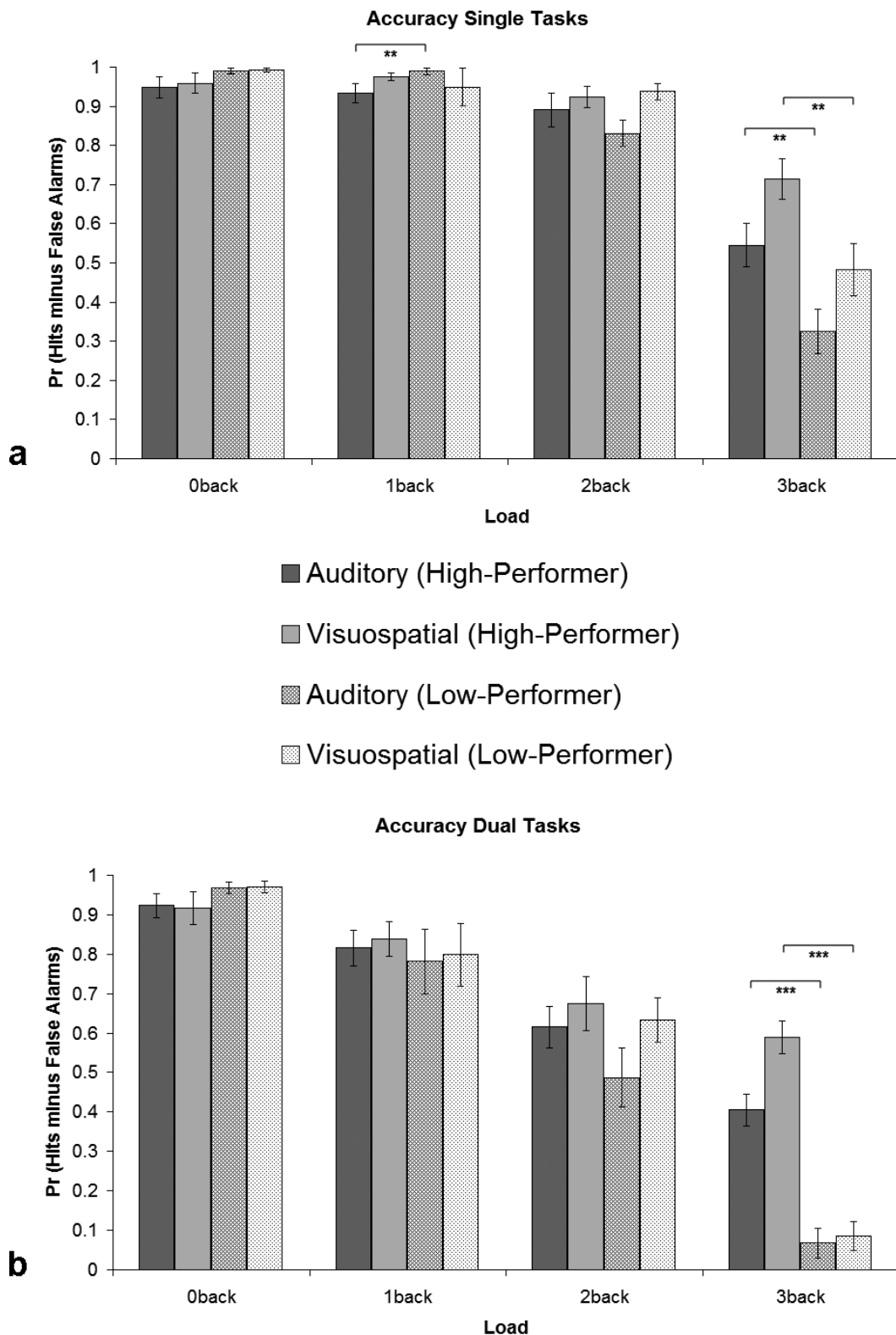
Figure 17. Means and SEM in the various levels of the n-back task (a: Single tasks; b: Dual tasks) for each performance group as differentiated by performance scores in the 3-back dual tasks (Experiment 5). Low-performing group: $N = 11$; high-performing group: $N = 13$. Significant group differences are indicated (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Accuracy Single Tasks

a

■ Auditory (High-Performer)

□ Visuospatial (High-Performer)

▨ Auditory (Low-Performer)

▢ Visuospatial (Low-Performer)

Accuracy Dual Tasks

b
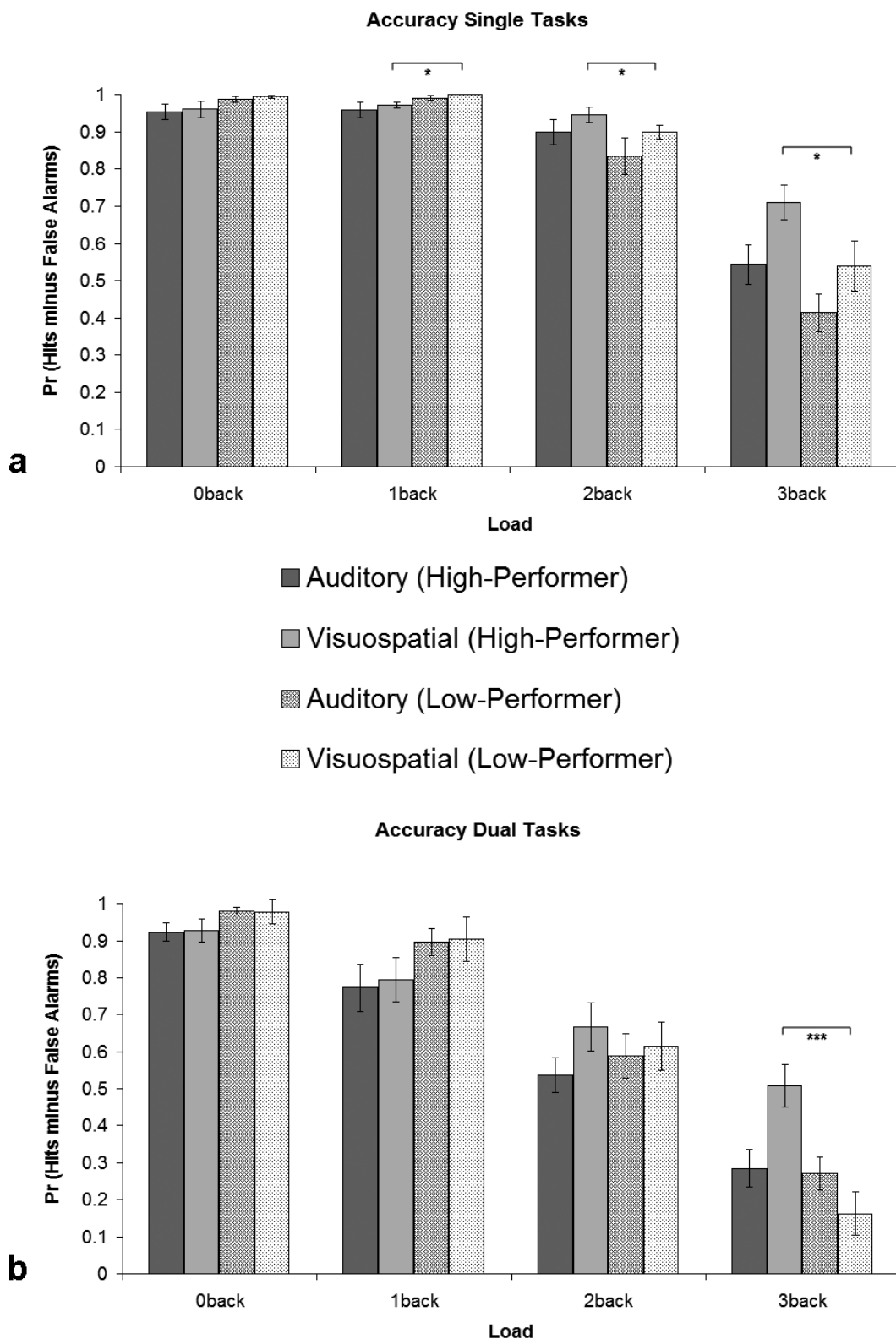
Figure 18. Means and SEM in the various levels of the n-back task (a: Single tasks; b: Dual tasks) for each performance group as differentiated by the APM-scores (Experiment 5). Low-performing group: $N = 13$; high-performing group: $N = 17$. Significant group differences are indicated (* $p < 0.05$; *** $p < 0.001$).

Experiment 1a. Correlating the different levels of the n-back task with the crystalline intelligence measures (i.e., with the WST), lower correlations were observed than in Experiment 5 (see Table 13). Again, the RTs in the n-back task were not significantly correlated with performance in the WST, apart from the visuospatial 3-back dual task ($r = 0.23$; $p < 0.05$). In terms of accuracy, the dual-tasks yielded larger correlations with the WST than the single tasks, which were, however, moderate, with the highest correlation for the auditory 2-back dual task ($r = 0.39$; $p < 0.01$).

Once more, two groups of participants were selected, based on their performance levels in the 3-back dual task. The sample consisted of 26 low-performing participants ($P_r < 0.17$; $M = 0.08$; $SD = 0.08$) and 25 high-performing participants ($P_r > 0.41$; $M = 0.50$; $SD = 0.10$), and again, significant differences between these two groups were observed also in other levels of the n-back task in terms of accuracy (*single tasks*: Visuospatial 3-back ($Z = -3.38$; $p < 0.001$); auditory 3-back ($Z = -3.62$; $p < 0.001$); *dual-tasks*: Visuospatial 1-back ($Z = -3.35$; $p < 0.001$), visuospatial 2-back ($Z = -3.16$; $p = 0.001$), visuospatial 3-back ($Z = -5.61$; $p < 0.001$), auditory 1-back ($Z = -2.93$; $p < 0.01$), auditory 2-back ($Z = -2.99$; $p < 0.01$), auditory 3-back ($Z = -5.43$; $p < 0.001$); all one-tailed; see Figure 19), while in RTs, the only significant differences were observed in the visuospatial 1-back dual task ($Z = -2.32$; $p = 0.01$; one-tailed) and in the visuospatial 2-back dual task ($Z = -1.85$; $p < 0.05$; one-tailed), where high-performing participants responded significantly faster. Conversely, the WST scores did not differ significantly between the two samples: The mean WST score for the low-performing participants was 33.35 ($SD = 4.47$), whereas it was 34.88 ($SD = 1.92$) for the high-performing participants ($Z = -1.32$; $p < $ ns).

Further, two groups of participants were differentiated, based on their performance in the WST: The high-performing group consisted of 31 participants (WST score > 36; $M = 36.94$; $SD = 1.03$), whereas the low-performing group consisted of 29 participants (WST score < 33; $M = 30.83$; $SD = 3.70$). No differences were observed at any level of the n-back task between these two groups of participants, neither in accuracy, nor in RTs.
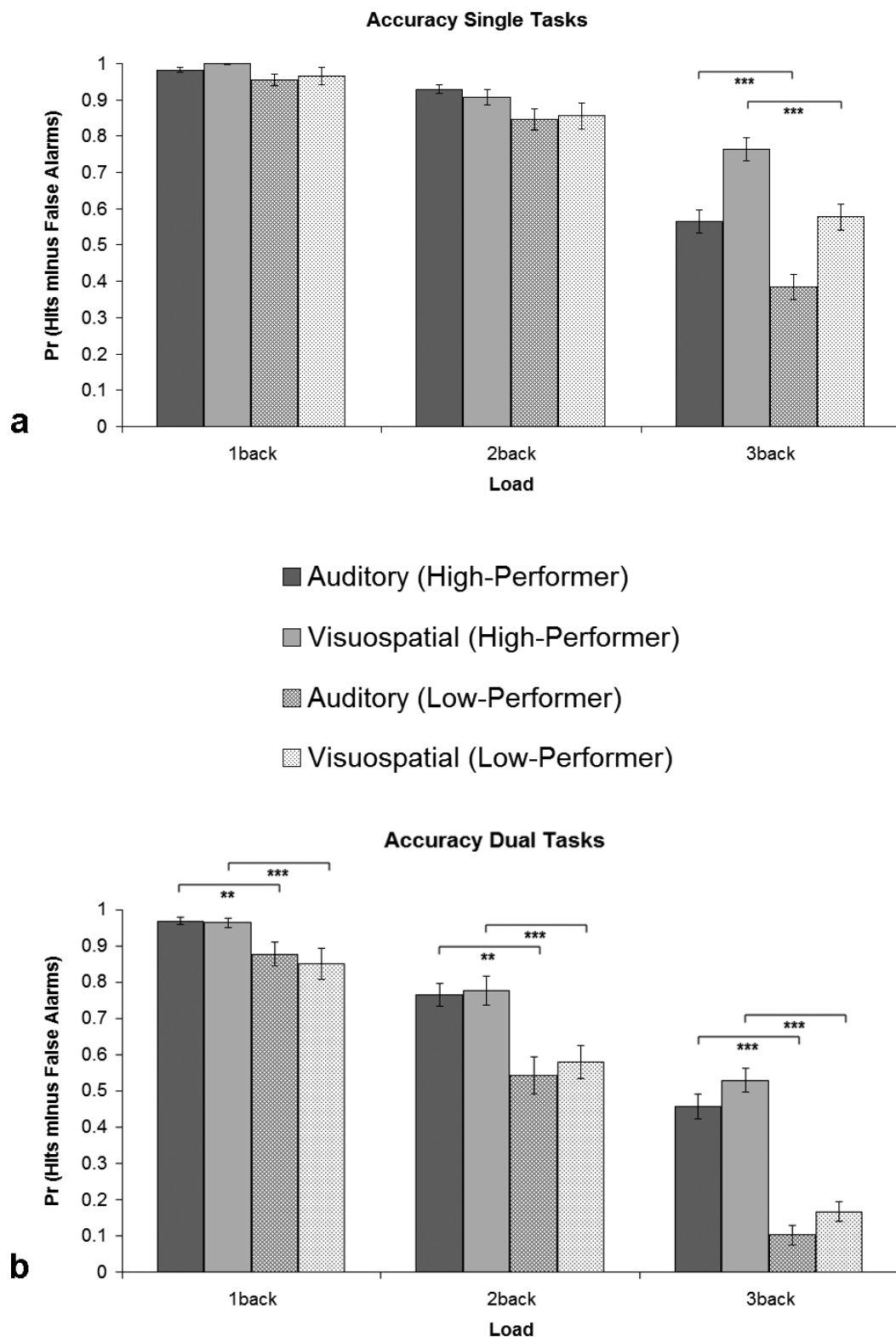
Figure 19. Means and SEM in the various levels of the n-back task (a: Single tasks; b: Dual tasks) for each performance group as differentiated by performance scores in the 3-back dual tasks (Experiment 1a). Low-performing group: $N = 26$; high-performing group: $N = 25$. Significant group differences are indicated (** $p < 0.01$; *** $p \leq 0.001$).

6.1.4. Discussion

In Experiment 5, some considerable correlations are observed between the n-back task and the intelligence measure representing g, with the largest correlation found between the 3-back visuospatial dual task and the APM ($r = 0.60$). The predicted increasing correlations between the APM and the increasing levels of the n-back tasks were not clear-cut, however, the only significant correlations are expressed in the 3-back tasks (except for the 3-back dual auditory task; $r = 0.12$; $p = $ ns). Especially in the dual tasks, no increase in correlation-coefficiens were observed between the APM and the n-back task in increasing load, suggesting that in the dual tasks, participants used different strategies in the various levels of load and also in the different modalities, which seem to relate differentially to g. The correlation coefficients in the visuospatial single tasks show a slight increase regarding the load and are comparable to, or, as in the 3-back task, slightly higher than the ones reported by Hockey and Geffen (2004). The absent or low correlations in the n-back tasks at lower levels of load (0- and 1-back tasks) are probably due to ceiling effects, therefore not differentiating between participants.

In contrast to my hypothesis, the dual-task versions of the n-back task did not yield higher correlations with the APM apart from the 3-back visuospatial dual task, therefore, no clear evidence that WMC as assessed with dual tasks is stronger related to g than single tasks, as reported by Fogarty and Stankov (1982), or Spilsbury (1992) can be provided with the tasks used here.

Nevertheless, with the extreme group approach, the association between the n-back task and the APM as indicated with the correlation approach could be strengthened: Those participants scoring high in the APM were also better in the (visuospatial) versions of the n-back task, especially at highest levels of load (3-back), and similarly, participants with better performance in the 3-back dual task scored higher in the APM. Therefore, it can be concluded that the n-back task at the 3-back level does very well differentiate between participants with high and low IQ and thus, seems to differentiate well between participants with high and low WMC.

The domain-free hypothesis could only be supported in the single tasks, where comparable correlations of either the visuospatial or the auditory n-back task with the APM were observed, especially in the 3-back conditions. However, in the dual-task conditions, only the visuospatial 3-back task was significantly correlated with the

APM, suggesting, that in the dual tasks, only the processes accounting for the performance in the visuospatial 3-back task are shared with the APM, which seem nevertheless substantial.

No evidence was found regarding the complexity hypothesis by Jensen in regard to RTs: Although the highest correlation was found between the 1-back auditory single task (an intermediate level in the view of Hockey & Geffen, 2004), this correlation was nevertheless moderate (and not significant), and also, the participants were clearly at ceiling performance in this task, making the interpretation difficult. In the dual tasks, there were no correlations, implying therefore, that there is more than 'mental speed' to the contribution of ability measures, at least as assessed here with the n-back task. I have already stated in Experiment 1, that there seems to be no speed-accuracy tradeoff in the various levels of load, and therefore suggesting that participants rather rely on accuracy and less on speed in performing these tasks.

Crystalline intelligence as tapped with the WST however, does only seem weakly relate to WMC, with the highest correlation in the auditory 2-back dual task, which is also consistent with findings by Hockey and Geffen (2004), reporting low correlations between the verbal-scale of a MAB and the visuospatial version of the n-back task as assessed with accuracy. However, they observed some correlations between the RTs and the verbal MAB, which was not present in my experiment, apart from one significant correlation between the 3-back visuospatial dual task and the WST ($r = 0.23$); suggesting again, that RTs does not contribute much to intelligence as assessed with the tasks I used here. Taking the extreme groups into account, no difference was observed in terms of the WST if they were differentiated on their n-back performance, and similarly, no differences in either level of the n-back task were observed, if groups were differentiated regarding their WST scores, neither in accuracy, nor in RTs. It seems therefore, that other, mediating factors might contribute to the correlations between the n-back task and the WST, maybe found more in the domain of g.

To summarize, despite the somehow mixed results, the findings of Experiment 5 and 1a provide further evidence that WMC and intelligence are truly related constructs. Larger correlations are observed if WMC is assessed with dual-tasks, and if intelligence is measured with a test relating to fluid intelligence, such as the APM. Based on these results, it can be concluded that the n-back task, especially when

applied as dual task at high levels of load, can well predict performance in other tasks representing higher order cognitive functions. This relationship is strengthened if the results of the extreme groups are taken into account, and can therefore also be regarded as a valid measure to estimate individual differences in WMC.

7. Are capacity limitations extendable?

7.1. Some basic reflections

There have been many reports about individuals being able to extend their memory capacity after excessive training, and systematic research goes back to Binet (1894) and earlier. The most prototypical task, in which improvements in WMC were demonstrated, is the classical memory span task (e.g. a digit span). With a considerable amount of practice, some participants were able to store up to 20 digits (Ericsson, 1988), while two participants studied by Ericsson and Chase (1982) could memorize more than 80 digits. However, their basic WMC remained the same as before, that is, their memorizing ability did not transfer to other tasks, such as remembering consonants instead of digits (Ericsson & Delaney, 1998), since their performance relied on very specific strategies and also primarily on associations already existing in their long-term memory (LTM). The fact that their acquired expertise did not generalize into another domain implies that WMC is constraint to domains, which can be optimized through domain-specific skills and knowledge.

Still, two complementary questions can be raised in regard to the above mentioned findings: First, can we train our basic capacities with these types of training? And second, do we rather adopt and adapt strategies in order to process these tasks with a higher degree of proficiency and/or efficiency with the capacity remaining unchanged? These two questions relate to the old debate of nature vs. nurture and are still widely discussed in the literature, the first again gaining importance through the huge body of research in genetics and their impact on cognition these days (see Goldberg & Weinberger, 2004, for a review). The historic view relating to the nature-assumption was held by Galton (1869; as cited in Ericsson & Lehmann, 1996), that 'individual differences reflect innate basic capacities that cannot be modified by training and practice' (p. 274). On the other end, relating to the nurture-assumption, there is considerable amount of evidence that expert performance, as demonstrated with chess players, is based on experience and knowledge, and usually results from years of practice in their domain (Chase & Simon, 1973; de Groot, 1946). However, these types of exclusive distinctions do not help to clarify underlying mechanisms in plasticity, learning, and expert performance, and it seems rather more appropriate to distinguish the underlying processes not in an

exclusive way, but in acknowledging, that they might well be qualitatively different but nevertheless interacting.

Since the effects of skills on memory performance have been demonstrated by many authors (see above), the question remains therefore, whether it is possible to alter the 'hardware', i.e., the basic abilities with deliberate training, that is, without the use of strategies relying on knowledge and/or LTM. In terms of intelligence-theories (see Chapter 5), the similar question can be posed, namely, whether it is possible to enhance the fluid component of intelligence. The crystalline component is per definition prone to accumulated knowledge, therefore not limited to a certain amount (provided that the individual has access to that knowledge) and is also preserved up to very old age (e.g. Park et al., 2002). Fluid intelligence on the other hand is much more regarded as fixed entity which seems, similarly as WMC, not modifiable with training (i.e., without the use of strategies), and, other than crystalline intelligence, shows developmental changes resembling an inverted u-shaped curve with the highest performance in young adulthood (for a review, see Lövdén & Lindenberger, 2005).

As mentioned earlier in the thesis, attentional deficits are among the most prevalent impairments in brain-damaged patients, estimated to be prevalent in 80% of the neuropsychological patient population (Van Zomeren, Brouwer, & Deelman, 1984). In clinical neuropsychology therefore, many attempts to 'retrain' these basic attention impairments have been made, also in order to improve other cognitive functions depending on attention. Thus, 'attentional training' is widely applied in neuropsychological rehabilitation settings. Still, a recent literature review by Leclercq and Sturm (2002) and also, an efficacy study by Sturm et al. (2003) involving a wide range of patients, end up with somehow mixed and controversial results: While in some studies, no improvements resulted from the attention training (e.g. Malec, Jones, Rao, & Stubbs, 1984; Middleton, Lambert, & Seggar, 1991; Ponsford & Kinsella, 1988), others report minor changes, however limited to the trained function (e.g. Gray & Robertson, 1989; Gray, Robertson, Pentland, & Anderson, 1992; Sturm, Dahmen, Hartje, & Willmes, 1983; Sturm & Willmes, 1991; Wood, 1986). Only a few studies observed some kind of generalization effect to other, non-trained functions (Ben-Yishay, Piasetzky, & Rattok, 1987; Lamberti, Wieneke, & Franke, 1988; Sohlberg & Mateer, 1987; Sturm, Willmes, Orgass, & Hartje, 1997). In sum, the results in this field are most inconsistent, which might be due to some important limitations: First,

there might be many more studies, finding no improvements and not being published because of these non-effects, and second, there is a lack in adequate control groups in most of these studies, and third and most limiting, as pointed out by Ponsford and Kinsella (1988), if spontaneous recovery and mere practice effects are controlled for, many of the above-mentioned improvements disappear. Further, it remains open, whether any of these effects persist after the training has abandoned, and also, treatment benefits on the patients' daily life have not been proved yet.

A large part of the existing training literature addresses the question, whether it is possible to remediate cognitive functions in schizophrenic patients. The interest in cognitive remediation in this patient group is comprehensible regarding the fact that cognitive impairment is a characteristic symptom in schizophrenia (e.g. Bellack, Gold, & Buchanan, 1999). The observed impairments mainly affect attentional processes, memory and executive functions (Braff, 1993; Calev, 1984; Green, 1998; Hoff et al., 1999; Rund, 1998). Although Pilling et al. (2002) concluded in their review that there are no beneficial effects in cognitive remediation of schizophrenic patients, there is now a whole body of literature that convincingly shows the opposite outcome (Bark et al., 2003; Bellack, Weinhardt, Gold, & Gearon, 2001; Fiszdon, Bryson, Wexler, & Bell, 2004; Medalia, Revheim, & Casey, 2002; Sartory, Zorn, Groetzinger, & Windgassen, 2005) with even evidence for long-lasting effects over several months (Fiszdon et al., 2004; Medalia et al., 2002; Wexler et al., 1997; Wykes et al., 2003). Besides showing specific improvements in the trained tasks, there are also studies demonstrating generalized training effects, i.e., improvement in cognitive tasks that were not specifically trained (e.g. Bell, Bryson, Greig, Corcoran, & Wexler, 2001; Bellack et al., 2001). Although these results are promising, it has to be stated that research on cognitive remediation in psychiatry is still young and some concerns, mainly on methodological basis, restrict the above-mentioned findings. It is not evident for example, by which types of tasks cognitive functions can be improved most efficiently. Further, the problem of adequate control groups is the same as in attention-training studies with brain-lesioned patients as mentioned above. In sum, more methodologically convincing studies must be conducted in order to allow firm differential conclusions about the beneficial effects of cognitive remediation in schizophrenic patients.

In this context, it seems therefore surprising, that some recent studies (e.g. Klingberg et al., 2005; Klingberg, Forssberg, & Westerberg, 2002; Posner & Rothbart, 2005) show that with a WM (a visuospatial span-board training) or computerized attention training, not only increased WMC was demonstrated, but even generalizing effects such as increased general fluid intelligence as assessed with the Raven Progressive Matrices (Raven, 1990). These patterns have been observed in children with Attention Deficit Hyperactivity Disorder (ADHD) (Klingberg et al., 2005; Klingberg et al., 2002; Experiment 1), very young healthy children (Posner & Rothbart, 2005), but also, with a very small healthy student sample (Klingberg et al., 2002; Experiment 2; Olesen et al., 2004). The attractive and common factor between these studies is, that the training is constructed in an adaptive way, that is, the better the person gets in the task, the more difficult it becomes, therefore always challenging the abilities of the participants and thus preventing boredom.

However, taking a closer look at the healthy student sample (Klingberg et al., 2002; Experiment 2), we discovered huge treatment effect sizes in the Raven Test, which were in the range of 4.79 (Cohen's d; our estimates from the available data; pre vs. post test). These effect sizes raised our interest, even that we had some reservations on the validity of these results: Looking at the individual learning curves for example, it appeared that most of the improvements in the WM task took place in the first three days of practice, and could therefore account for mere familiarization with the task. We had also concerns regarding the applied intelligence test, because it is possible that improvements are observable simply because of re-testing – would the same effects be observable in a non-treatment control group? Klingberg et al. (2002) did not use a control sample, for his student group; rather, they took the control children of the ADHD sample as control for the student sample, even though different versions of the Raven Test were used. Further, although a generalizing effect on *g* has been reported with an ADHD sample (Klingberg et al., 2002; Experiment 1), this generalizing effect of the WM training on the IQ is less clear in a later study with healthy students (Olesen et al., 2004). In this study, the effects of a visuospatial WM training on the BOLD response were studied, and in the pre and post sessions, among other measures (Stroop and a visuospatial WM task), also Raven's APM were applied. However, the exact results of this task were not reported, and similarly, nothing was reported about the performance of the control group, therefore leaving

the question open, whether an improvement could also be possible with mere re-testing.

In spite of the shortcomings of these studies, the potential of such WM training is fascinating and the possible impact in theoretical aspects is challenging. Would it be really possible to improve the WMC after some days of training, without teaching the use of strategies? Does this training really 'boost' the IQ as reported by these studies mentioned above? What would that mean for the theories of fluid intelligence and WMC? It is somehow interesting that these questions are not appropriately discussed in the above-mentioned studies, although they seem essential for the existing literature of intelligence and WMC.

In order to clarify these questions we[9] decided to conduct a similar study in order to replicate the results with an even more challenging task, i.e., with an adaptive dual task, but also to improve the experimental conditions from a methodological point of view, for example with the use of adequate control groups.

---

[9] Myself and Martin Buschkuehl, with the help of some students in terms of the experimental training, as well as two students during a research project.

7.2. Extending capacity limitations by means of adaptive dual-task training

7.2.1 Does dual-task training really work? Experiment 6

In this experiment, the effects of a 10-day training of an adaptive version of an n-back dual task were studied. The adaptive version should be very directly depending on the actual performance of the participant: Not being too easy, but also not too difficult; always providing a sense of achievement in the participant in order to keep the motivation high. Comparing pre and post measures, effects on the task itself were evaluated, but also effects on other WM measures, and on a measure of fluid intelligence.

The following hypotheses were tested: First of all, a typical logarithmic learning curve was expected, that is, that after a few sessions, no significant improvement in the n-back task would be possible anymore, i.e., the curve should become asymptotic in the sense of a plateau being reached after a few training sessions, and thus representing a fixed upper bound on performance. Secondly, compared to the baseline-measurements, the post session should show an improvement in the n-back tasks at higher levels of load (at lower levels of load, i.e., in the 0- and 1-back tasks, ceiling levels were expected, and therefore, no improvements were expected there). The third hypothesis concerned other WM measures also applied in pre and post sessions: As further measure of WM, the digit-span task (forward and backwards) from the HAWIE-R (Tewes, 1991) was chosen in order to represent WM. In the forward version mainly the phonological loop, i.e., the verbal subcomponent is assumed to be tapped. In the backward version however, additional and more 'executive', i.e., controlled processes are assumed to take place, since apart of passively storing the items, they have to be mentally manipulated in order to recall them in reverse order, thus being in line with the processes attributed to the central executive of the WM model as proposed by Baddeley (1986). This mental reversal process might be related to (visual) imagery processes, and indeed, factor analysis indicated that both, verbal and visual processes contribute to performance in the reversed digit span (Larrabee & Kane, 1986). The dual-task version of the n-back task, with material being manipulated in both modalities is therefore assumed to share some common processes with the digit-span task. Consequently, we expected improvements in both measures of the digit-span task (forward and backwards) in

response to the n-back training, but still, more improvements were expected in the reversed version because we supposed that with the n-back task not solely storing capacity is trained, but also manipulation processes, which should be more expressed in the backwards version. Since several studies (e.g. Kane et al., 2005; Kyllonen & Christal, 1990) and Chapter 5 demonstrated substantial shared variance between the concepts of WMC and fluid intelligence, the findings reported by Klingberg et al. (2005; 2002), as well as Posner and Rothbart (2005), i.e., the effects of WM training generalizing on fluid intelligence, seem comprehensible, despite some reservations stated above: If improvements in one task would be observed, improvements should be also observable in another, strongly related, task. As a last hypothesis we therefore expected an increase in performance in the APM as a measure of fluid intelligence, providing that we would obtain significant improvements in the n-back WM task.

The results of the n-back training group in the post-test session were further compared to those of a control group without training (control group I), and with those of a second control group with unspecific RT training (control group II). Control group II was chosen in order to control for possible unspecific task improvements (e.g. daily training environment, familiarization with the testing situation and the experimenter) and mere speed, which is sometimes seen as basis for other *g*-related tasks (e.g. Salthouse, 1996; Salthouse et al., 2003). Therefore, comparing the performance in the n-back training group to control group I and control group II, mainly differences were expected in the n-back tasks at higher levels of load. This difference should be more pronounced in control group I than in control group II. The group with the RT training instead should react faster, mainly in the 0- and 1-back tasks, and also, some other enhancements regarding WM measures, solely because of speed improvements are expected, which would be in line with hypotheses by Salthouse (Salthouse, 1996; Salthouse et al., 2003). For the control group I, no significant improvements are expected on any measure.

7.2.2. Method

Participants. The sample consisted of 23 participants, mainly undergraduate students. There were 12 women and 11 men; the mean age was 24.48 (*SD*: 3.96).

Apparatus. Task administration was computerized for the n-back and run on a Microsoft Windows based computer programmed with E-prime (Psychology Software Tools, Pittsburgh, PA). The other tasks, i.e., the APM and the digit-span task were administrated as paper and pencil tests.

Task Design and Procedure. The participants were divided in three groups: There was one experimental group (N=8), which accomplished an adaptive n-back training (see below) for 10 days, preceded and followed by pre and post tests. Further, there was a first control group (N = 8), only attending pre and post tests, and a second control group (N=7), which also completed 10 training sessions as well as pre and post tests, but with 'unspecific' and non-adaptive RT trainings (see below) as intervention, with the same duration as the experimental group. All participants were first familiarized with the n-back task as described in the other experiments. The two training groups, i.e. the experimental group and control group II first completed two days of training before the "pre test" was administered at day three. This was implemented in order to rule out the possibility that any improvements in the post-test measures would result due to a habituation to the testing situation and towards the investigators.

The training sessions took place on 10 consecutive days, with a two-day break after 5 days. One training session lasted approximately 30 minutes each, and consisted of 20 runs. Within a run, the same n-back task was presented (e.g. a 2-back task) for $20 + n$ trials (n was depending on the value of the n-back task). Depending on the performance of the participants, either the same task load was repeated for the next $20 + n$ trials, or a new level of load was administered, according to the following criteria: If the participant had a performance rate of 90% and more (calculated separate for each modality, i.e., visuospatial *and* auditory and therefore requesting the participants' to divide their attention equally between the two modalities), the following run was increased in load ($n +1$). If the participants' performance was in the range of 70-90%, n remained at the same level of load, whereas the level was decreased by n-1, if the participants' performance was below 70%. If the participant made no responses at all (for example in choosing a very conservative strategy, i.e.,

making no false alarms, but also no hits), the participant dropped one level of n. The same was true for the reverse strategy (if the participant responded to each item). The exact percent criterion values varied depending on the value of n. The changes in load depending on the performance of the participants were controlled by the program, and instructions for participants, indicating the level of load they had to undertake, were determined automatically and presented before each new run of $20 + n$ trials. Before each new run, participants were given the opportunity to take a short brake. On the first three days, participants started their training with a 1-back task, but from session 4 to session 10, they started with a 2-back task in order to make sure that participants reached their optimal level of performance soon enough, but also to provide a short 'warm-up' period to get familiar with the task.

Responses had to be made with the left index finger on key 'A' on the keyboard in regard to the visuospatial material, whereas key 'L' had to be pressed with the right index finger in regard to the auditory material.

The 'unspecific' RT tasks consisted of four different tasks, with the intention to improve speed only. All were choice-RT or simple RT tasks, presented in the visual modality: A visuospatial 0-back task, an 'inhibition of return' task, a visual search task, and a simple RT task, each lasting approximately 5 minutes, resulting in a comparable amount of training-time in each session as the n-back training group.

For pre and post sessions, a baseline n-back task was administered, the digit-span task (backward and forward), and the Raven's APM.

Material.

N-back training: The same task material was used for all the 10 training sessions. The stimuli were the same as in the experiments before and consisted of blue squares presented at 8 different locations in the visuospatial condition, and 8 auditorily presented consonants. The stimuli were presented for 500 ms, followed by an ISI of 2500ms. In 10 trials of each run, no responses were required, in 4 trials a response had to be made to the visuospatial material only, in 4 trials to the auditory material only, and in 2 trials a response had to be made to both types of material simultaneously. The amount of targets, i.e., of requested responses was therefore always the same (i.e., 6 per 20 trials per modality; not present in the first n trials), but they were presented at random positions, determined by the program.

RT training: Four computerized tasks were administered in each session.

Task 1 was a simple single visuospatial 0-back task as described in experiment 2, i.e., participants had to press key 'L' as fast as possible with their left index finger, if the square was presented in the upper left corner. No responses were required, if the square appeared in the other seven possible locations. The ratio of targets and non-targets was 1:3 and the distribution of the targets was determined randomly.

Task 2 was an 'inhibition of return task' in courtesy and adapted from Pascal Wurtz' experiments: Participants viewed a fixation cross in the middle of the screen, and had to react as fast as possible adequately with their right index finger, if a cross appeared on the right on the screen, and accordingly with their left index finger, if the cross appeared on the left. No response was required if a circle instead of a cross was shown in order to force participants to constantly pay attention. The left-right distribution was determined randomly. Further, squares surrounding the crosses were presented as cues, either valid, i.e., being consistent with the later shown crosses, or invalid, i.e., shown on the other side. The distribution of the cues was also determined randomly.

Task 3 comprised a classical feature detection task as described by Treisman (Treisman, 1964; Treisman & Gelade, 1980): An array of brown and green letters (x and n) was presented on the screen randomly changing in size (from 5, 15 or 30 letters) and participants had to decide as fast as possible, whether a blue-coloured letter was additionally included in the array. The presence or absence of the letter was determined randomly. Participants received feedback after each trial, whether their response was correct and how long their RT was.

Task 4 was a simple RT task: 8 stimuli (pictures of flowers) were presented in a circle on the screen and participants had to click as quickly as possible with the mouse on the accordant stimulus, if it was covered with a red square. Feedback concerning the RTs was provided in order to make sure that the participants constantly try to react fast enough.

Pre and post sessions.

N-back task. In the baseline condition, the same tasks were used as described in Experiments 1 and 5, but with some additions: Six levels of difficulty (0- to 5-back) were administered as dual task.

Digit-span task. The digit span as forward and backwards version as used in the HAWIE-R (Tewes, 1991) was administered as a measure of WM. Participants listened to sequences of digits, which had to be verbally repeated, either forward or backwards. Per amount of digits two trials were presented (starting with 2 digits in the backwards span and with 3 digits in the forward span). The span was determined as sequence, where both trials had been perfectly recalled by the participant.

Raven Advanced Progressive Matrices (APM): The APM as described in Experiment 5 were used, however, because no parallel-versions of the APM exist, half of the test (18 items; all even numbers) was administered in the pre session and the other half (18 items; all odd numbers) was administered in the post session. The amount of time provided to complete the test was therefore reduced from 40 to 20 minutes; a procedure also reported in the study by Klingberg et al. (2002).

Analysis. Accuracy and RTs were evaluated for all three groups for the n-back baseline task (dual-task condition) as well as the accuracies in the training task for the n-back training group. Additionally, the digit-span score as well as the APM score were determined. The RTs in the training sessions in the RT-training group were not analyzed. Repeated measures ANOVAS with session (pre vs. post) as within-subject factor and treatment group (experimental group, control groups I and II) as between factor were calculated for each measure, in order to assess, whether there were group and session effects, and most importantly, whether the groups showed differential effects, i.e., whether the experimental group showed effects in the post session, which could be attributed to the n-back. In case of significant differences, non-parametric tests were administered for all tasks where the Kolmogorov-Smirnov test indicated non-normal distributions; in the other cases, *t*-tests were applied. According to the hypotheses, two-tailed significance tests were used for between-group comparisons in the pre session (where no group differences were expected) and one-tailed tests for the post-test comparisons. Comparing the pre and post sessions within-groups, one-tailed tests were used for the experimental group and control group II (RT training), and two-tailed tests for control group I (no training).

Further, the effect sizes (*ES*), i.e., the effect of the treatment was calculated according to Cohen (1988) for each measure and separately for both groups of participants:

$$ES = \frac{\mu_1 - \mu_2}{\sigma_{pooled}}, \text{ whereas } \sigma_{pooled} = \sqrt{\frac{(\sigma_1^2 + \sigma_2^2)}{2}} \qquad (7\text{-}1)$$

As we were using a repeated design in comparing the pre and post measures for each group, it has been suggested to use a corrected *ES* formula for repeated measurements (e.g. Bortz, 1999) also considering reliability (Meier, 1999):

$$ES_{corr} = \frac{\mu_1 - \mu_2}{\sigma_{pooled}\sqrt{1 - r_{\mu1\mu2}}} \qquad (7\text{-}2)$$

However, Dunlap, Cortina, Vaslow, and Burke (1996) argued that if the pooled standard deviation is corrected for the amount of correlation, i.e., of the reliability between the measures, then the *ES* estimate will be an overestimate of the actual *ES* in the case of a high correlation, and, on the other hand resulting in a smaller effect in the case of low reliability because of more error variance (Meier, 1999). Consequently, both *ES* measures will be evaluated and reported here, considering the differential impact of reliability in the various measures, but also bearing in mind, that the sample size is very small. The test-retest reliability was calculated as Pearson's Product Moment Correlation (*r*) between the pre and the post session for each group separately.

7.2.3. Results

Generally, participants of the *n-back training group* showed an impressing learning curve over the 10 training sessions: Starting at a mean performance level of *n* = 2.30 (*SD*: 0.43) across the whole first session (i.e., of the last 17 runs out of 20; excluding the first three runs in order to make sure to include only these trials, in which participants were approximately at their actual training level), they reached a mean performance level of *n* = 4.16 (*SD*: 0.64) in the last session (corresponding to an accuracy score of > 90% in both modalities). The difference between the first and the last training session was highly significant ($t(7) = - 8.00$; $p < 0.001$; one-tailed). The mean performance increase for all participants over the training sessions is depicted in Figure 20.
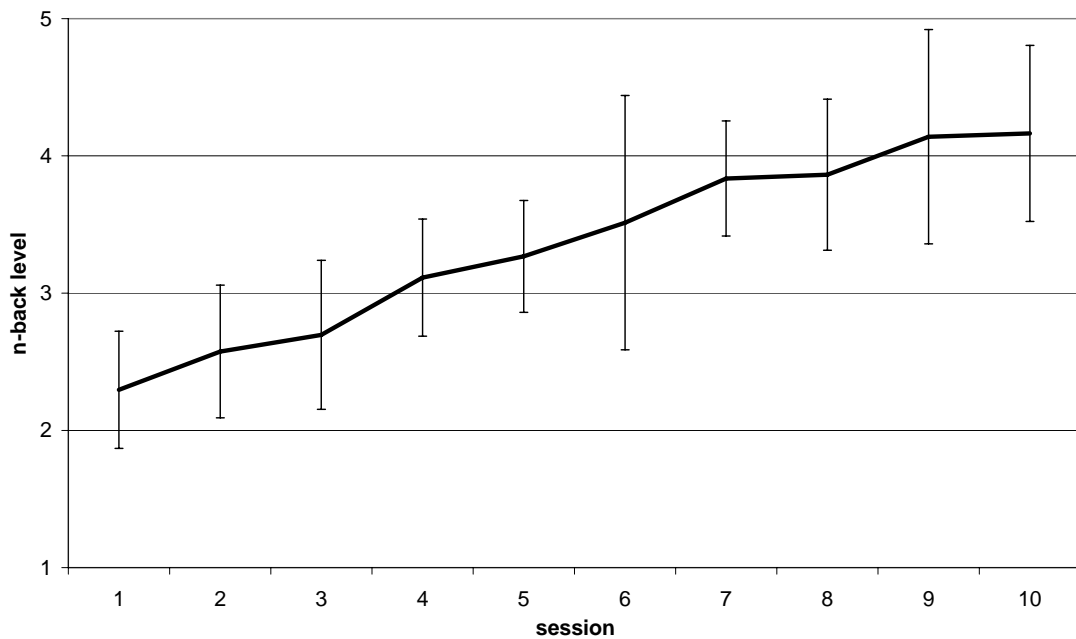
Figure 20. Mean performance level for every training session as well as the corresponding standard deviations (*N* = 8). Each data point represents the mean performance level (*n*) of the last 17 runs (out of 20 runs per session).

Even after session 10, the data were better represented with a linear function ($R^2 = 0.85$; $F(1,168) = 948.82$; $p < 0.001$), than with a logarithmic function ($R^2 = 0.71$; $F(1,168) = 417.40$; $p < 0.001$), suggesting that participants had not yet reached their performance limit at the last session and that with more training, even higher performance levels would be obtained (see Figure 21).
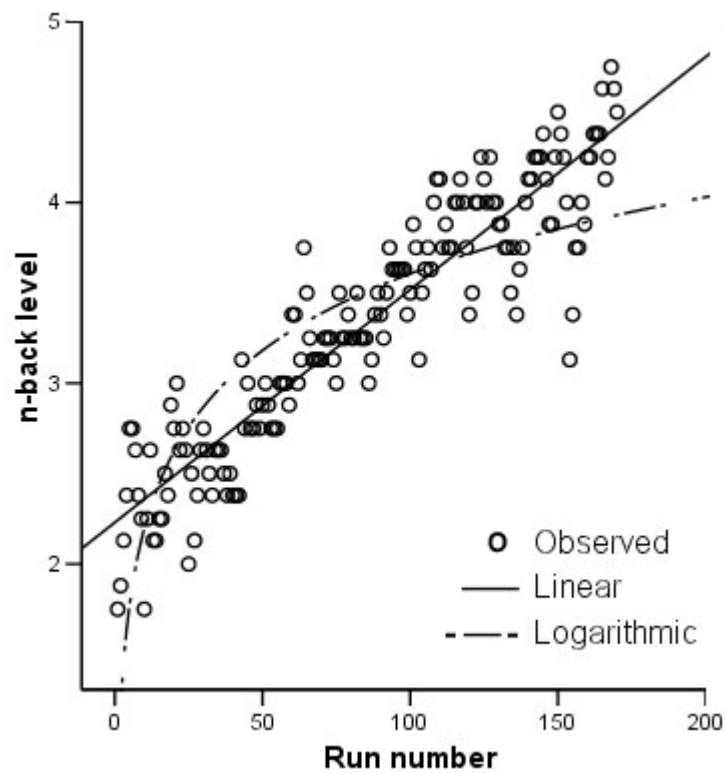
Figure 21. Mean performance level for each run and superimposed hypothetical equations tested in the model, i.e., the linear and the logarithmic function.
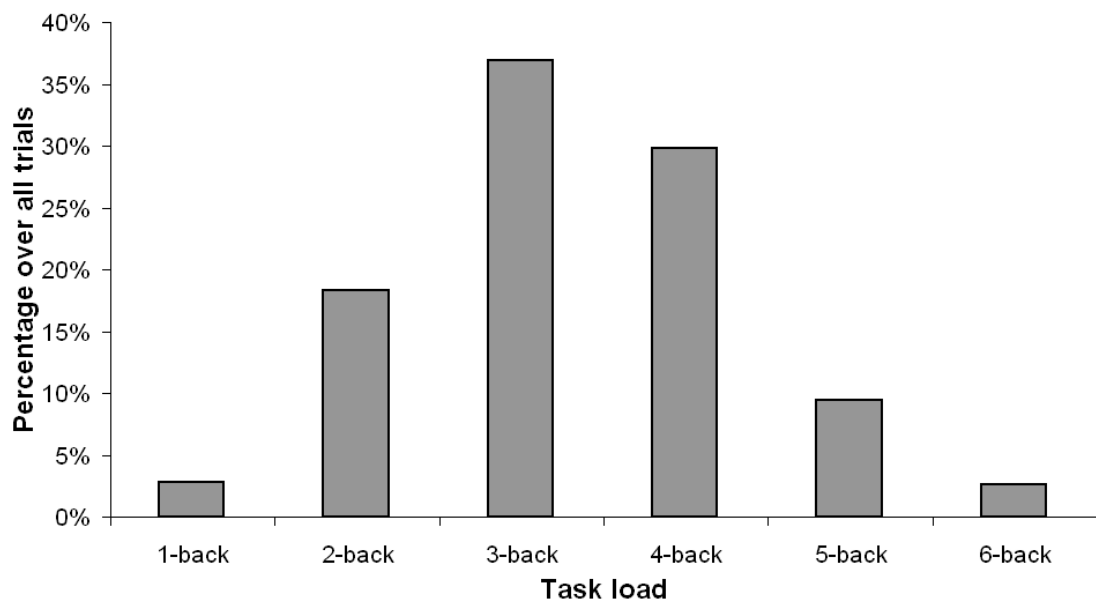
Figure 22. Proportion of n-back levels that participants were performing over the 10 training sessions (i.e., comprising the last 17 runs of each session for all participants; *N* = 8). The first three runs are excluded for reasons already described in Figure 20.

Looking at the various levels of load obtained over the 10 training sessions, it can be seen on Figure 22 that the participants spent most of the time in practicing the 3-back task (36.91%), but also the 4-back task was performed in almost one third of the time (29.85%).

The repeated measures ANOVA for all measures with session as within-subject factor and treatment group as between factor yielded only one main effect for group in the visuospatial 3-back task (accuracy; $F(2,23) = 4.36$; $p < 0.05$), where the experimental group performed significantly better than the control group I ($t(14) = 2.89$; $p < 0.05$; two-tailed). A similar trend for the visuospatial 4-back task (accuracy; $F(2,23) = 3.18$; $p = 0.06$) was observed, and further, there was a trend for a group-effect in the auditory 0-back task (RT; $F(2,23) = 2.72$; $p = 0.09$). In all other tasks, no main effects for group were observed, indicating therefore, that the three groups were comparable. There were some significant main effects for session; in two of them, a significant improvement from the baseline to the post session was observed, namely, in the digit-span task (*forward*: $F(1,23) = 13.38$; $p < 0.01$; $t(22) = -3.29$; $p < 0.01$; *backwards*: $F(1,23) = 17.20$; $p < 0.001$; $t(22) = -4.26$; $p < 0.001$; two-tailed), with a trend in the same direction for the visuospatial 4-back task ($F(1,23) = 3.18$; $p = 0.07$; accuracy) and for the auditory 0-back task ($F(1,23) = 3.30$; $p = 0.08$; RT). In the auditory 1-back task (accuracy), the participants showed a significant performance decrease from the pre to the post session ($F(1,23) = 5.43$; $p < 0.05$; $t(22) = 2.45$; $p < 0.05$; two-tailed). Several interactions between session and treatment group revealed that the experimental group showed significantly better performance in the post session, while the two control groups showed no improvements. This was true especially at higher levels of load in the n-back accuracy measures (*auditory 3-back*: $F(2,23) = 5.71$; $p < 0.05$; $t(7) = -2.84$; $p = 0.01$; *visuospatial 5-back*: $F(2,23) = 4.22$; $p < 0.05$; $t(7) = -3.94$; $p < 0.01$; one-tailed), but also in the digit-span task (*forward*: $F(2,23) = 3.91$; $p < 0.05$; $t(7) = -4.41$; $p < 0.01$). In two measures, the experimental group outperformed the control group I in the post session (*visuospatial 5-back task* (accuracy): $t(14) = 1.97$; $p < 0.05$; forward *digit-span task*: $t(14) = 2.26$; $p < 0.05$; both one-tailed). Conversely, the control group II (RT training) responded faster in the visuospatial 1-back task in the post session ($F(2,23) = 5.08$; $p < 0.05$; $t(6) = 4.45$; $p < 0.01$; one-tailed). In this task, control group II was also significantly faster than the experimental group in the post session ($t(13) = 2.92$; $p < 0.01$; one-tailed). In the

auditory 1-back task, this group was faster than both other groups in the post session (*control II vs. experimental*: $t(8.57) = 2.64$; $p < 0.05$; *control II vs. control I*: $t(13) = 3.08$; $p < 0.01$; one-tailed), whereas the experimental group was significantly slower in the post session compared to the baseline ($t(7) = -2.01$; $p < 0.05$; one-tailed). Also in the visuospatial 0-back task (accuracy), the experimental group showed a significant performance decrease from the baseline to the post session ($F(2,23) = 3.67$; $p < 0.05$; $t(7) = 2.06$; $p < 0.05$; one-tailed), while the other groups did not show any difference. For all other measures, the interactions did not reach significance. The control group I without training did not show any significant difference between the two sessions in any measure. All descriptive data, as well as significant differences within and between groups are indicated on Table 14.

The *ES* speak for the validity of the two training conditions, the n-back and the RT training: In terms of accuracy, the largest *ES* (Cohen's d) are observed in the experimental group, especially at higher levels of load (e.g. *ES* = 0.99 in the visuospatial 5-back task), and the same applies to the digit-span task (*forward*: *ES* = 1.41; *backwards*: *ES* = 0.92). Concerning RTs, the largest improvements (i.e., faster responses) are observed in control group II, for example in the auditory 0-back condition (*ES* = 1.11). Control group I (without training) usually shows the smallest *ES* of the three groups (see Table 14). As also shown on Table 14, the *reliability* measures considerably vary between the three groups, reflecting the variability of the small sample size, but which could also be attributed to the effect of the different treatments between the groups. Comparing the two types of *ES*, one considering reliability, and the other without taking these measures into account, it can be well observed, that the corrected *ES* increases with the value of *r*.

Table 14. Descriptive data, effect sizes, and Test-Retest Reliability (*r*) for the different levels of load, as well as for the digit span and the APM for the two test sessions shown for each group of participants.

Dual n-back task		Pre Test		Post Test		Absolute Mean Difference (pre vs. post)	Test-Retest Rel. (*r*)	*ES* (Cohen's d)	*ES* (corrected)
		M	(SD)	M	(SD)				
RTs (ms)									
0-back auditory	Experimental group	624	(106)	625	(188) b	1	0.11	0.01	0.01
	Control group I	680	(217)	620	(181) c	60	0.94	0.30	1.24
	Control group II	542	(108)	**442**	(68) *	100	0.53	1.11	1.61
1-back auditory	Experimental group	868	(278)	**1106**	(489) *,b	238	0.75	0.60	1.20
	Control group I	921	(269)	894	(181) c	27	0.73	0.12	0.22
	Control group II	723	(224)	624	(155)	99	0.71	0.51	0.95
2-back auditory	Experimental group	924	(219)	1033	(513)	109	0.07	0.28	0.29
	Control group I	1085	(293)	1154	(214) c	69	0.54	0.27	0.40
	Control group II	1026	(336)	**794**	(258) *	232	0.61	0.77	1.25
3-back auditory	Experimental group	1175	(292)	1050	(505)	125	-0.05	0.30	0.30
	Control group I	1260	(421)	1327	(572)	67	0.64	0.13	0.22
	Control group II	1129	(418)	940	(319)	189	0.19	0.51	0.57
4-back auditory	Experimental group	1097	(288)	1167	(528)	70	0.17	0.16	0.18
	Control group I	1128	(407)	920	(183)	208	0.87	0.66	1.82
	Control group II	880	(289)	893	(299)	13	0.29	0.04	0.05
5-back auditory	Experimental group	1139	(472)	1049	(503)	90	0.41	0.18	0.24
	Control group I	1040	(505)	1034	(533)	6	0.75	0.01	0.02
	Control group II	1210	(430)	1132	(444)	78	-0.19	0.18	0.16
0-back visuospatial	Experimental group	577	(136)	774	(372) b	197	0.33	0.70	0.86
	Control group I	617	(152)	567	(165)	50	0.89	0.32	0.93
	Control group II	509	(102)	478	(77)	31	0.64	0.34	0.57
1-back visuospatial	Experimental group	808	(223)	939	(268) b	131	0.65	0.53	0.89
	Control group I	840	(357)	791	(265)	49	0.91	0.16	0.52
	Control group II	721	(206)	**596**	(168) **	125	0.94	0.67	2.74
2-back visuospatial	Experimental group	957	(420)	1066	(684)	109	0.84	0.19	0.49
	Control group I	1088	(273) c	973	(371)	115	0.42	0.35	0.46
	Control group II	692	(273)	857	(349)	165	0.10	0.53	0.55
3-back visuospatial	Experimental group	1081	(316) b	1071	(512)	10	0.59	0.02	0.04
	Control group I	1164	(455) c	1026	(310)	138	0.62	0.35	0.57
	Control group II	728	(222)	868	(345)	140	0.61	0.48	0.78
4-back visuospatial	Experimental group	1134	(342)	1062	(399)	72	0.37	0.19	0.24
	Control group I	1153	(528)	901	(570)	252	0.66	0.46	0.78
	Control group II	886	(236)	897	(327)	11	0.76	0.04	0.08
5-back visuospatial	Experimental group	1004	(351)	1175	(622)	171	0.57	0.34	0.51
	Control group I	1024	(458)	915	(199)	109	0.16	0.31	0.34
	Control group II	852	(337)	911	(427)	59	0.90	0.15	0.48
Accuracies (Pᵣ)									
0-back auditory	Experimental group	0.75	(0.05)	0.66	(0.25)	0.09	-0.60	0.50	0.39
	Control group I	0.83	(0.13)	0.77	(0.11)	0.06	-0.13	0.50	0.47
	Control group II	0.75	(0.14)	0.78	(0.17)	0.03	0.20	0.19	0.21
1-back auditory	Experimental group	0.70	(0.12)	0.6	(0.15)	0.10	0.35	0.74	0.91
	Control group I	0.75	(0.05)	0.69	(0.10)	0.06	0.55	0.76	1.14
	Control group II	0.77	(0.07)	0.72	(0.15)	0.05	-0.16	0.43	0.40
2-back auditory	Experimental group	0.52	(0.12)	0.47	(0.19)	0.05	0.02	0.31	0.32
	Control group I	0.43	(0.10)	0.36	(0.18) c	0.07	-0.12	0.48	0.45
	Control group II	0.42	(0.18)	0.58	(0.23)	0.16	0.41	0.77	1.01
3-back auditory	Experimental group	0.25	(0.14)	**0.46**	(0.17) *	0.21	0.09	1.35	1.42
	Control group I	0.39	(0.23)	0.29	(0.30)	0.10	0.74	0.37	0.74
	Control group II	0.37	(0.23)	0.32	(0.16)	0.05	0.68	0.25	0.44
4-back auditory	Experimental group	0.29	(0.14)	0.41	(0.23) b	0.12	0.20	0.63	0.70
	Control group I	0.27	(0.21)	0.3	(0.14)	0.03	0.53	0.17	0.24
	Control group II	0.19	(0.13)	0.19	(0.13)	0.00	0.56	0.00	0.00
5-back auditory	Experimental group	0.20	(0.15)	0.27	(0.11) a	0.07	0.45	0.53	0.72
	Control group I	0.15	(0.16)	0.16	(0.07)	0.01	-0.35	0.08	0.07
	Control group II	0.19	(0.16)	0.17	(0.14)	0.02	0.65	0.13	0.22

Table 14. (Continued).

Dual n-back task (Pr)		Pre Test M	(SD)	Post Test M	(SD)		Absolute Mean Difference (pre vs. post)	Test-Retest Rel. (*r*)	*ES* (Cohen's d)	*ES* (corrected)
0-back visuospatial	Experimental group	0.94	*(0.09)*	**0.82**	*(0.18)*	*	0.12	0.34	0.84	1.04
	Control group I	0.83	*(0.17)*	0.90	*(0.06)*		0.07	0.09	0.55	0.58
	Control group II	0.90	*(0.11)*	0.85	*(0.10)*		0.05	0.82	0.48	1.11
1-back visuospatial	Experimental group	0.84	*(0.17)*	0.69	*(0.31)*		0.15	-0.08	0.60	0.58
	Control group I	0.81	*(0.15)*	0.85	*(0.14)*	c	0.04	0.87	0.28	0.78
	Control group II	0.70	*(0.32)*	0.68	*(0.15)*		0.02	0.54	0.08	0.12
2-back visuospatial	Experimental group	0.69	*(0.26)*	0.64	*(0.21)*		0.05	0.52	0.21	0.31
	Control group I	0.62	*(0.15)*	0.66	*(0.18)*		0.04	0.78	0.24	0.52
	Control group II	0.61	*(0.22)*	0.60	*(0.27)*		0.01	0.75	0.04	0.08
3-back visuospatial	Experimental group	0.65	*(0.29)* a	0.63	*(0.32)* a		0.02	0.66	0.07	0.11
	Control group I	0.25	*(0.24)*	0.31	*(0.29)*		0.06	0.34	0.23	0.28
	Control group II	0.46	*(0.25)*	0.36	*(0.26)*		0.10	0.76	0.39	0.80
4-back visuospatial	Experimental group	0.35	*(0.19)* a	0.42	*(0.12)* a		0.07	0.06	0.44	0.45
	Control group I	0.15	*(0.18)*	0.20	*(0.20)*		0.05	0.64	0.39	0.65
	Control group II	0.21	*(0.24)*	0.32	*(0.22)*		0.11	0.63	0.48	0.79
5-back visuospatial	Experimental group	0.13	*(0.20)*	**0.31**	*(0.16)*	**,a	0.18	0.76	0.99	2.03
	Control group I	0.20	*(0.14)*	0.16	*(0.13)*		0.04	0.79	0.30	0.65
	Control group II	0.22	*(0.19)*	0.18	*(0.15)*		0.04	-0.14	0.23	0.22
Digit Span										
Forward	Experimental group	7.00	*(1.60)*	**9.50**	*(1.93)*	**,a	2.50	0.60	1.41	2.23
	Control group I	7.25	*(2.60)*	7.50	*(1.60)*		0.25	0.75	0.12	0.23
	Control group II	7.57	*(2.99)*	8.57	*(1.99)*		1.00	0.89	0.39	1.18
Backwards	Experimental group	7.38	*(1.85)*	**9.13**	*(1.96)*	***	1.75	0.89	0.92	2.81
	Control group I	6.88	*(2.03)*	8.00	*(2.27)*		1.12	0.68	0.52	0.92
	Control group II	7.43	*(1.62)*	8.29	*(2.21)*		0.86	0.70	0.44	0.82
Intelligence Test (Raven APM)										
	Experimental group	13.75	*(2.92)*	14.00	*(2.93)*		0.25	0.84	0.09	0.21
	Control group I	11.75	*(2.25)*	12.13	*(2.59)*		0.38	0.45	0.16	0.21
	Control group II	14.43	*(2.88)*	12.86	*(1.46)*		1.57	0.29	0.69	0.82

Note. N = 23; RTs are reported as an average of *median* ms. *M* = mean; *SD* = standard deviation; *ES* = effect size.

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.01$ (difference between pre and post session; one-tailed for experimental and control group II; two-tailed for control group I); significant differences between pre to post session are further indicated with bold font.

a: Significant difference between experimental and control group I (no training)   ⎤   ($p < 0.05$; pre session:
b: Significant difference between experimental and control group II (RT training) ⎬ Two-tailed; post
c: Significant difference between control group I and control group II   ⎦ session: One-tailed)

7.2.4. Discussion

As Figure 20 shows, an impressive improvement in the dual n-back task was obtained with the n-back training – the n-back level of 4.17 was not expected and the expected approximation to an asymptote was not reached, suggesting that WM training with the n-back task is indeed possible and would possibly show even more improvements after 10 sessions of training. Also, the improvement in the n-back tasks at higher levels of load and in the other WM task (i.e., in the digit-span task) was large and much more pronounced in the experimental group than in either control group as indicated with the significant interactions. All groups however showed improvements in the digit-span task, indicating that some effects can be obtained with mere replications of a task and with simple and unspecific RT trainings only. Yet, the differential effects in the experimental group, showing larger improvements in the forward digit-span version in addition to the improvements in the high-load n-back tasks suggest a generalizing effect going beyond the practiced tasks. Since no other tasks were used, the question remains open, whether this generalizing effect is limited to the WM domain, or whether it would manifest itself in other cognitive domains, such as attention, memory and executive functions.

Inconsistent to Klingberg's results (2002), no improvements in the APM were observed in any group. There are several explanations for this finding, mainly on methodological basis: First, in the pre- as well as in the post-test sessions, some participants were at a ceiling level, especially in the experimental group. For this reason, no improvements were possible anyway. Another, even more important reason could lie in the low reliability of the test, as applied here and in Klingberg et al. (2002): Dividing a test per se already lowers reliability (see Chapter 3.2.1.1.), and it seems that the APM looses far too much reliability if only half of the test is used: Over the three groups, Pearson's Product Moment Correlation between the two halves was $r = 0.58$ ($p < 0.01$), and moreover, it was also very different in the three groups (experimental group: $r = 0.84$; $p = 0.01$; control group I: $r = 0.45$; $p = $ ns; control group II: $r = 0.29$; $p = $ ns). Despite the low number of participants, this low and unsystematic reliability measures are striking, and suggest that the procedure of dividing the APM into two parts and using it as two parallel test versions, as it was successfully done and reported by Klingberg et al. (2002) was obviously not appropriate for our sample and certainly contribute to the results found here. The

hypothesis, that there would be observable improvements in the APM (if the WM training would work) could therefore not be verified or falsified. Thus, with a better and more reliable measure of fluid intelligence, a more specific conclusion should be possible. Further, the linear trend in the n-back training observed here led to the next experiment, in which the n-back training comprised twice the training sessions as in this experiment in order to investigate whether limitations in this task would be visible later in the training. Also, a more complete test-battery was selected, containing more tasks and also control tasks in order to be able to differentiate between cognitive domains being sensitive to WM training, and others being not.

The exact procedure and the results of this study will be described as following in Experiment 7.

7.3. Generalizing effects of adaptive dual-task training

7.3.1. Can we obtain generalizing and differential effects on WM and ability measures

with dual-task training? Experiment 7

As stated before, this study was conducted in order to replicate and extend the findings of Experiment 6: I was primarily interested to see whether an asymptotic curve regarding performance would be reached after nearly twice of the training sessions used in Experiment 6, and further, whether generalized and differential effects on various cognitive tasks could be obtained with this training. Therefore, more tasks were included compared to Experiment 6, covering many aspects of WM (i.e., verbal tasks, visuospatial tasks), executive functions, as well as control tasks not used in Experiment 6 in order to investigate whether the WM training has a selective effect on tasks which are related to the concept of WM and executive functions with no effect on these control tasks. With respect to fluid intelligence, a more appropriate task than the APM, i.e., the 'Bochumer Matrizentest' (BOMAT; Hossiep, Turck, & Hasella, 1999) was used, which has the advantage that full parallel-versions are available and that the task was explicitly developed in order not to yield ceiling effects in student samples. The experiment was carried out together with Martin Buschkuehl and Daniela Blaser; the latter writing her Master thesis on the topic.

7.3.2. Method

Participants. 16 participants agreed to take part in the study; 8 participants, i.e., 4 women and 4 men (mean age: 25.5; *SD*: 2.27) comprised the experimental group, completing an n-back training over four weeks. One participant of this group had to be excluded from the data analyses due to too many missing data in the training sessions[10]. In the control group, there were 8 participants; 6 women and 2 men (mean age: 25.13; *SD*: 1.13). The two groups were comparable in education, being mostly students and recruited by Daniela Blaser.

Material and Procedure. The n-back training (material and procedure) was the same as in Experiment 6, with the exception that 19 instead of 10 sessions were performed by the experimental group (the 20th session was cancelled because of a public holiday).

The baseline and the post sessions (duration: 60 minutes each) took place on three consecutive days, in advance of the training period and on two days after the 19th training session (60 and 90 minutes), and there, various WM and other cognitive and ability tasks were assessed:

WM tasks.

− The *digit-span task* (forward and backwards) from the HAWIE-R (Tewes, 1991) as described in Experiment 6 as representing mainly the phonological loop (however, see Chapter 6.2.2. for a discussion concerning the backwards version).

− A *visuospatial-span task* (Vs2) was administered in order to represent the visuospatial sketchpad. A computerized (E-Prime; Psychology Software Tools, Pittsburgh, PA) and visuospatial analogue to the digit-span task resembling the Corsi Block Span (Milner, 1971; Schellig, 1993) was developed by Martin Buschkuehl and me beforehand. In this task, participants are presented with a 4x4 grid on the computer screen in which a sequence of positions is shown with a circle-shaped dot. The dot remains at each position for 750 ms, with an ISI of 1000 ms, and the positions are determined

---

[10] This was the only participant undertaking his training sessions by himself at home without fixed appointments, which resulted in varying training-times and forgetting to undertake the training on several occasions.

randomly. Participants have to repeat this sequence either forward or backwards by clicking with the mouse on the required positions on the grid remaining on the screen. As in the digit-span task, there were always two trials per amount of positions, and the experiment is terminated, if the participant is not able to reproduce the sequence in either of the two trials.

The visuospatial span is determined as the highest level, in which the participant is able to remember both trials of the sequence.

– For the central executive system, a partly computerized version of the *RST* (programmed with E-Prime; Psychology Software Tools, Pittsburgh, PA) with the same material as described in Experiment 1 was administered; however, some improvements were made due to further experiences with this task in other experiments. The main difference to the RST version in Experiment 1 was that the sentences which the participants had to read aloud were presented on the computer screen and that the 'yes'/'no' responses had to be made on the keyboard (1 for 'yes', and 0 for 'no'). As soon as the response was made, the next sentence was presented, but if the participant took too long to decide (implying extensive rehearsal processes), a blank computer-screen followed after a few seconds (8.1 sec for short sentences and 8.6 for long sentences; based on the mean decision time in an earlier experiment with $N = 53$) and remained there until the decision was made. After 2, 3, 4, 5, or 6 sentences, participants were required to recall the last words of each sentence, which the experimenter was protocolling.

Executive and speed-tasks.

– A shortened version of the *Stroop-task* (Bäumler, 1985) was used: In each session (pre and post), only three lists out of nine original lists were presented: The first list required participants to read names of colours (red, blue, yellow, green), the second to indicate the colours of bars, and finally in the actual interference condition, the colour of words printed in a colour different from the colour it actually names (e.g. the word 'red' printed in green) had to be reported as fast as possible. This measure seems to be a highly reliable assessment of the ability to inhibit overlearned answers to simple tasks (i.e., reading the name of the colour instead of saying the printed colour), and is

commonly taken as prototype measure of executive functions (e.g. Miyake et al., 2000; Salthouse et al., 2003). Therefore, treatment and group differences are expected mainly on the interference condition.

- *Task-Switching*: This task was adapted from Kray and Lindenberger (2000) administered with a computer using E-Prime (Psychology Software Tools, Pittsburgh, PA). Participants had to conduct two different tasks: In task A they had to decide whether the first two digits of four presented digits were larger (by pressing key 'x') or smaller (by pressing key 'm') than the last two digits, whereas in task B they had to decide whether two or more of the four presented digits were identical (by pressing key 'm' if two digits were identical, otherwise by pressing key 'x'). After having accomplished task A for 32 trials and task B for further 32 trials, participants were given 8 exercise trials in a block were they had to switch continuously between task A and task B: After having accomplished the same task for two times, participants were required to switch to the other task, resulting in the following sequence: AABBAABB. 4 identical blocks with 32 trials were administered, each starting with task A requiring the subjects to switch tasks like in the exercise. The participants had to keep track of the tasks themselves, that is, no indication of the task to be done was given[11]. The stimuli consisted of 192 different 4-digit numbers which were presented in the middle of the screen in white forecolour on a black background. Every answer was equally often possible for both tasks (i.e., pressing key 'x' or 'm' in task A or task B). The exercise block consisted of 8 unique 4-digit numbers. All numbers were presented randomly. After having pressed either key 'x' or key 'm' a response stimulus interval (RSI) of 200ms followed (i.e., a blank screen). Each new block was preceded with a written instruction. Participants' performance was assessed in evaluating RTs separately in non-switch trials and in switch-trials. Switch-costs, as indicator of task-switching ability results in subtracting the switch RTs from the non-switch RTs. Since task switching is sometimes seen as executive function (e.g. Miyake et al.(2000); however, see Meiran, Chorev,

---

[11] Some earlier experiments with this tasks executed by Martin Buschkuehl not reported here showed, that despite that participants had to remember the sequence by heart, rarely errors were made, indicating that this condition did not pose too much memory load on them.

& Sapir (2000)), we expected the experimental group having less switch-costs in the post session than the control group.

- The *digit-symbol test* from the HAWIE-R (Tewes, 1991) was taken as a complex speed-measure. 90 squares in several rows were presented, each paired with a randomly assigned number from 1 to 9. Above these rows, keys were depicted, pairing each number with a different nonsense symbol. Participants had to fill in the blank squares sequentially with the paired symbol as fast as possible for 90 seconds, and the score represents the number of correctly filled squares. As stated above, this task should represent a complex speed measure, but according to Lezak (2004), as a test of 'psychomotor performance' it should be 'relatively unaffected by intellectual prowess, memory, or learning' (p. 368) since most of the variance is explained through copy speed alone (Joy, Fein, Kaplan, & Freedman, 2000; Kreiner & Ryan, 2001) and also, incidental memory should contribute to performance in this test (Lezak et al., 2004). Commonly observed age differences in this test can therefore be best explained in motor slowing (Lezak et al., 2004). Following this line of arguments, no improvements in response to the WM training should occur in this task as well as no difference between groups.

Cognitive abilities.

- *Explicit memory* was assessed with word-list learning: Participants were presented with 20 standardized words derived from the word lists as used in the NAI (Oswald & Fleischmann, 1995) (one presentation only; 20 words were presented in the pre session and 20 other words in the post session). Participants had to recall as many words as possible, first, immediately after the presentation, and second, in a late recall condition at the end of the session (approximately 30 minutes later). We did not expect any group differences in this test in the first place; however, it is possible that with the experimental group becoming more efficient in organizing material, there might be some slight improvements in this group in the post test.

- *Fluid intelligence* was assessed with the BOMAT (Hossiep, Turck, & Hasella, 2001) using version A for the pre test, and version B for the post test. Further, in order to keep the testing-sessions short enough, the amount of time to

complete the test was limited to 30 minutes instead of 70 minutes (10 minutes for the practice items, and 20 minutes for the test). Following the arguments in Experiment 6, we expected improvements in the experimental group for this test comparing the scores in the pre and the post session.

Implicit memory.

A visual priming paradigm was used in which line drawings of various concrete pictures were presented on a computer screen. The pictures were derived from the International Picture Naming Project (IPNP; Szekely et al., 2005) and matched according to naming times and name agreement (Bates et al., 2003). The pictures were not presented at once, but built up slowly and pixelwise, and participants had to press the spacebar on the keyboard as soon as they thought they had recognized the picture. After pressing the spacebar, the screen remained blank until the participant gave his answer according to the picture verbally to the experimenter. The experimenter recorded the answer in pressing the mouse-button in regard to the answer (left: Correct answer; right: Wrong answer) and as soon as the mouse-button was pressed, the whole picture was presented at once for 2 s with a following pause of 1250 ms completing the trial. After that, a new picture was built up. 60 pictures were presented this way. The last 40 pictures included the 20 pictures presented initially, which were presented as 'old' pictures resulting in totally 60 trials with 40 different pictures. Priming was then determined as difference between the naming times of the first presented 20 pictures and the pictures presented the second time (only RTs of correct answers were included in the analysis). This difference was then transformed into a relative value of priming according to the following formula: % Priming = $(RT_{new}-RT_{old})/RT_{new}$ x 100. Further, RTs in response to the first 20 pictures ('new') were assessed as basic speed-component. In the pre and the post session, different pictures were presented. This task served as further control paradigm in the sense that no effects of training is expected in this task since priming is regarded as mainly automatic process (Perrig, Wippich, & Perrig-Chiello, 1993), which should not be altered by the use of explicit strategies or memory processes and should therefore not show performance changes in response to the training.

Finally, the *baseline n-back* was administered as described in Experiment 6 (0- to 5-back; dual task).

Analysis. As in Experiment 6, the mean level of the n-back training was assessed for each participant and each training session. Repeated measures ANOVAS with session (pre vs. post) as within-subject factor and treatment group as between factor and corresponding post-hoc tests were calculated as in Experiment 6, and similarly, *ES* and Test-Retest Reliability measures were assessed.

7.3.3. Results

In general, the results show that WMC and/or the processing efficiency as trained with the n-back task can be improved and the results of Experiment 6 could be well replicated. The general course of the training is depicted in Figure 23: The longer duration of the training did still not yield a performance-curve towards an asymptotic level as expected, yielding a mean performance level of 5.17 in the last session. All participants reached an n-back level of 7 at least in one session, and one participant even reached an n-back level of 9, implying continuously remembering 18 items. The proportion of obtained n-back levels are shown on Figure 24.
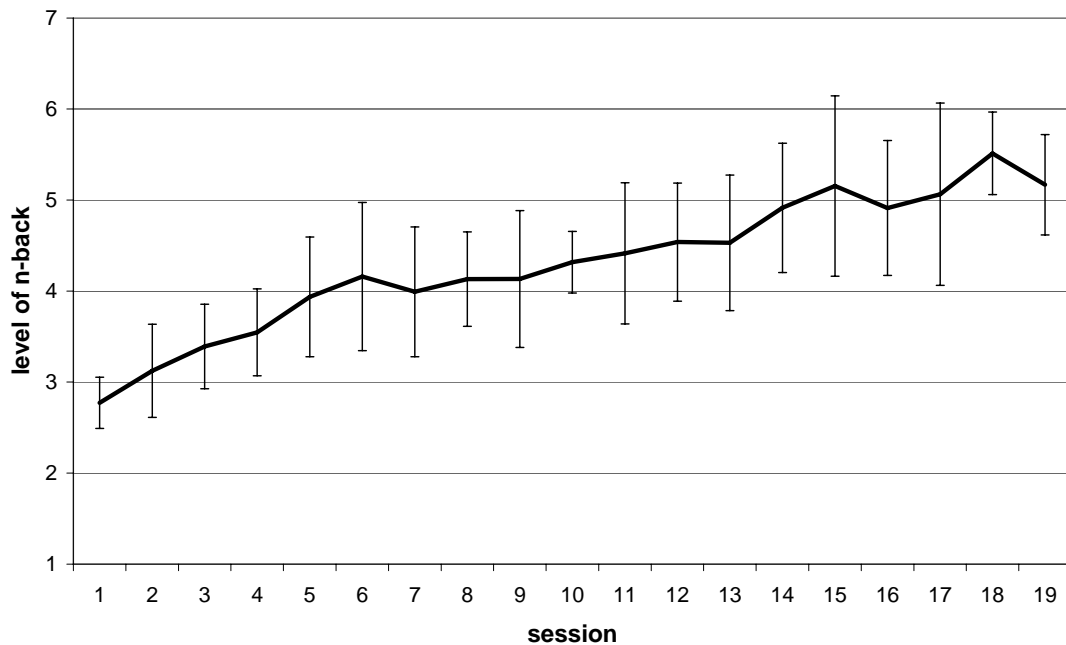
Figure 23. Mean performance level for every training session as well as the corresponding standard deviations ($N = 7$). Each data point represents the mean performance level (n) of the last 17 runs. The first three runs were excluded in order to make sure to include only the trials, in which participants were approximately at their actual training level.

Compared to Experiment 6 Figure 24 shows that participants spent now most of the time in practicing the 4-back task (31.62%) compared to the 3-back task in Experiment 6, and that the 5-back task was also practiced in almost one third of the time (27.24%).
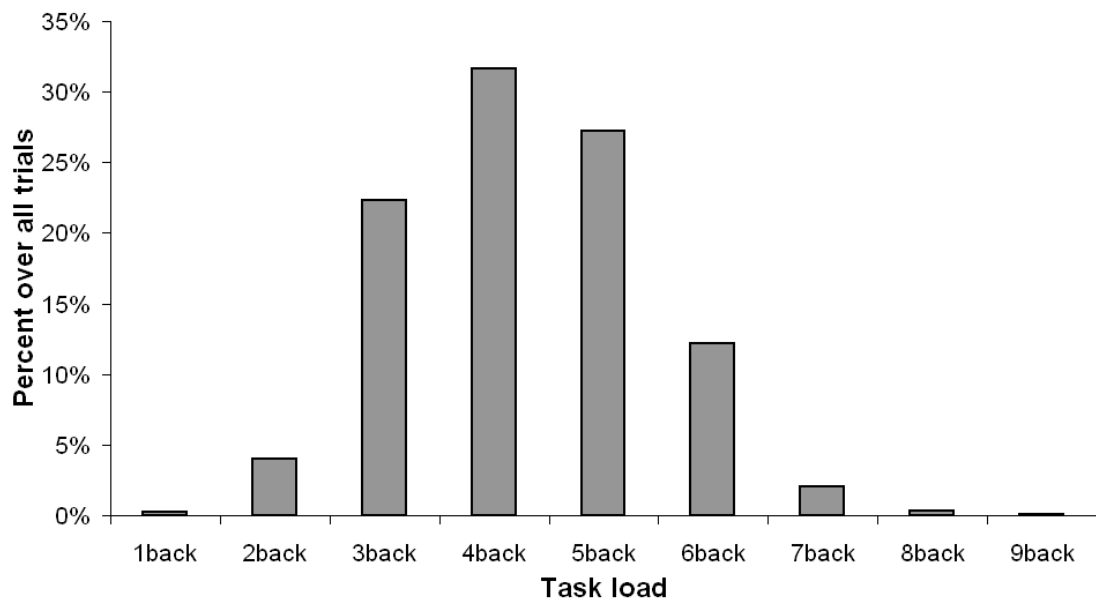
Figure 24. Proportion of n-back levels that participants were performing over the 19 training sessions (i.e., comprising the last 17 runs of each session for all participants; $N = 7$). The first three runs are excluded for reasons already described in Figure 23.

Even with 19 training sessions, the trend of linearity remained highly significant ($R^2 = 0.79$; $F(1,321) = 1226.16$; $p < 0.001$), however the logarithmic function also explains considerable amount of variance ($R^2 = 0.71$; $F(1, 321) = 781.46$; $p < 0.001$); see Figure 25.
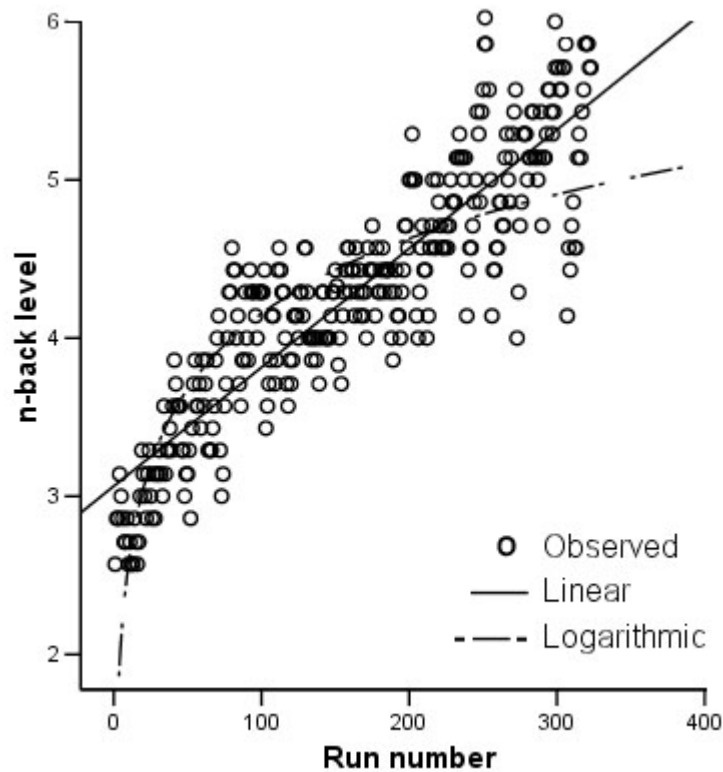
Figure 25. Mean performance level for each run and superimposed hypothetical equations tested in the model, i.e., the linear and the logarithmic function (see text for statistical values).

The results of the repeated measures ANOVA are shown in Table 15. To summarize, there were a few main effects of group, mainly in RTs in the n-back task (2- and 3-back; auditory and visuospatial) and on general speed (i.e. picture naming), where the experimental group responded faster in all tasks. The experimental group had also higher scores in the 3- and 4-back visuospatial tasks in accuracy. But apart of these main effects, the two groups were comparable. There were also many main effects of session, almost exclusively in the predicted direction of an improvement in performance from the pre to the post session. The only exception is the 0-back visuospatial task, where both groups made more errors in the post session. The main effect for session reaches significance mainly because of the experimental group, as indicated by the observed interactions: The experimental group showed significantly larger improvements in the post session than the control, suggesting that the treatment, i.e. the dual-task training had the intended effect. This differential effect was most pronounced in the n-back tasks at higher levels of load (RTs and accuracy), but also in executive measures, such as the digit span backwards and the RST. This

interaction was also present in ability measures, i.e. in the explicit memory (delayed recall) and in the intelligence test. There was also a speed measure, i.e. the Stroop colour naming task, where the experimental group was significantly faster in the post session than the control. There was one exception where, unexpectedly, the reverse was true: In the digit-symbol test, the control completed more items in the post session than the experimental group. The results of the post-hoc comparisons are reported in Table 16, where also all means and standard deviations, as well as the *ES* and reliability measures for the pre and post sessions are indicated separately for each group.

Table 15. Results for the 2-way repeated measures ANOVA (main effects and interaction) with session (pre vs. post) as within-factor and treatment group (experimental vs. control) as between-factor.

Dual n-back tasks	Group		Session		Group x Session	
	F	p	F	p	F	p
RTs (ms)						
0-back auditory	0.45	0.513	0.41	0.535	0.85	0.374
1-back auditory	3.68	0.077	2.64	0.128	0.05	0.835
2-back auditory	5.50	0.036*	4.49	0.054	2.41	0.145
3-back auditory	8.80	0.011*	1.26	0.282	4.08	0.065
4-back auditory	1.82	0.201	6.62	0.023*	6.13	0.028*
5-back auditory	0.08	0.777	7.98	0.015*	12.19	0.004**
0-back visuospatial	0.27	0.613	0.10	0.756	0.02	0.879
1-back visuospatial	3.85	0.073	4.19	0.063	1.60	0.230
2-back visuospatial	4.90	0.045*	38.36	0.000***	3.00	0.107
3-back visuospatial	12.93	0.004**	4.73	0.050*	6.67	0.024*
4-back visuospatial	1.89	0.192	2.08	0.173	10.74	0.006**
5-back visuospatial	0.52	0.484	20.30	0.001**	4.57	0.052
Accuracies (P_r)						
0-back auditory	0.14	0.718	0.73	0.408	0.22	0.648
1-back auditory	0.04	0.850	1.52	0.239	0.01	0.911
2-back auditory	0.12	0.731	5.39	0.037*	15.42	0.002**
3-back auditory	1.92	0.190	16.08	0.001***	20.49	0.001***
4-back auditory	0.64	0.440	11.79	0.004**	4.96	0.044*
5-back auditory	2.31	0.153	4.80	0.047*	16.23	0.001***
0-back visuospatial	1.31	0.273	5.70	0.033*	1.43	0.252
1-back visuospatial	2.56	0.134	0.06	0.813	1.06	0.322
2-back visuospatial	0.00	0.995	0.30	0.596	0.20	0.664
3-back visuospatial	9.96	0.008**	5.89	0.031*	15.94	0.002**
4-back visuospatial	8.38	0.013*	13.41	0.003**	19.70	0.001***
5-back visuospatial	4.48	0.054	15.07	0.002**	34.25	0.000***
Digit Span						
Forward	1.37	0.263	9.63	0.008**	0.30	0.593
Backwards	0.01	0.932	0.77	0.396	9.50	0.009**
Vs2						
Forward	0.54	0.475	0.01	0.926	2.01	0.180
Backwards	0.12	0.738	0.05	0.820	0.72	0.410
RST						
Span	0.40	0.540	15.92	0.002**	6.54	0.024*
Task Switching (ms)						
Non-switch	0.38	0.548	16.69	0.001***	1.74	0.210
Switch	0.33	0.575	5.48	0.036*	1.19	0.294
Switch-Costs	0.17	0.686	0.65	0.433	0.39	0.545
Stroop (s)						
Reading	0.20	0.660	0.38	0.549	1.23	0.288
Colour naming	1.62	0.225	15.21	0.002**	10.71	0.006**
Interference	0.82	0.381	1.62	0.225	1.84	0.198
Digit Symbol Test (Number of items)						
	1.96	0.185	4.17	0.062	5.29	0.039*
Priming						
General Speed (Picture naming)	9.27	0.009**	1.40	0.259	0.07	0.802
Priming (%)	0.30	0.591	3.07	0.103	0.01	0.941
Explicit Memory (Word list; Items recalled)						
Immediate recall	0.05	0.832	5.98	0.029*	4.21	0.061
Late Recall	0.00	0.969	10.05	0.007**	10.05	0.007**
Intelligence Test (BOMAT)						
	0.08	0.780	33.96	0.000***	5.20	0.040*

Note: Vs2: Visuospatial Span; RST: Reading Span Task. $df = 1$ for the main effects, and $df = 2$ for the interaction. Experimental group: $N = 7$; Control group: $N = 8$. *$p \leq 0.05$; **$p \leq 0.01$; *$p \leq 0.001$. The results for the post-hoc comparisons are reported on the next table (Table 16).

Table 16. Descriptive data, effect sizes and test-retest reliability (Pearson's Product Moment Correlation; *r*) for the various measures in the two test sessions shown for each group of participants.

Dual n-back tasks		Pre Test M	(SD)	Post Test M	(SD)		Absolute Mean Difference (pre vs. post)	Test-Retest Rel. (*r*)	ES (Cohen's d)	ES (corrected)
RTs (ms)										
0-back auditory	Experimental group	756	(160)	694	(98)		62	0.10	0.47	0.49
	Control group	766	(168)	777	(172)		11	0.73	0.06	0.12
1-back auditory	Experimental group	922	(196)	817	(275)		105	0.36	0.44	0.55
	Control group	1122	(213)	1041	(269)		81	0.79	0.33	0.73
2-back auditory	Experimental group	1166	(297)	**755**	(266)	*,a	411	-0.01	1.46	1.45
	Control group	1280	(246)	1216	(431)		64	0.17	0.18	0.20
3-back auditory	Experimental group	1133	(351)	**736**	(177)	**,a	397	0.55	1.43	2.14
	Control group	1360	(519)	1474	(432)		114	0.19	0.24	0.27
4-back auditory	Experimental group	1340	(493)	**750**	(287)	*,a	590	0.02	1.46	1.48
	Control group	1260	(387)	1249	(308)		11	0.59	0.03	0.05
5-back auditory	Experimental group	1331	(524)	**786**	(298)	**,a	545	0.61	1.28	2.04
	Control group	1117	(335)	1140	(296)		23	0.85	0.07	0.19
0-back visuospatial	Experimental group	640	(126)	645	(141)		5	0.32	0.04	0.05
	Control group	679	(205)	694	(205)		15	0.94	0.07	0.30
1-back visuospatial	Experimental group	890	(239)	680	(142)	a	210	-0.13	1.07	1.01
	Control group	992	(199)	942	(212)		50	0.68	0.24	0.43
2-back visuospatial	Experimental group	1096	(241)	**678**	(145)	**,a	418	0.32	2.10	2.55
	Control group	1258	(325)	**1023**	(216)	**	235	0.89	0.85	2.57
3-back visuospatial	Experimental group	988	(200) a	**672**	(149)	**,a	316	0.40	1.79	2.30
	Control group	1351	(362)	1272	(390)		79	0.10	0.21	0.22
4-back visuospatial	Experimental group	1320	(578)	**679**	(131)	**,a	641	0.43	1.53	2.02
	Control group	1092	(394)	1341	(381)		249	0.11	0.64	0.68
5-back visuospatial	Experimental group	1319	(453)	**674**	(116)	**,a	645	0.51	1.95	2.78
	Control group	1229	(465)	999	(307)		230	0.67	0.58	1.02
Accuracies (P$_r$)										
0-back auditory	Experimental group	0.81	(0.12)	0.82	(0.1)		0.01	0.35	0.09	0.11
	Control group	0.77	(0.12)	0.82	(0.12)		0.05	0.02	0.42	0.42
1-back auditory	Experimental group	0.72	(0.11)	0.66	(0.28)		0.06	0.54	0.28	0.42
	Control group	0.71	(0.06)	0.65	(0.09)		0.06	-0.62	0.78	0.62
2-back auditory	Experimental group	0.36	(0.15) a	**0.64**	(0.13)	***,a	0.28	0.53	1.99	2.90
	Control group	0.52	(0.13)	0.44	(0.14)		0.08	-0.12	0.59	0.56
3-back auditory	Experimental group	0.24	(0.18)	**0.71**	(0.14)	***,a	0.47	-0.12	2.91	2.76
	Control group	0.38	(0.22)	0.36	(0.17)		0.02	0.60	0.10	0.16
4-back auditory	Experimental group	0.23	(0.09)	**0.50**	(0.17)	**	0.27	-0.28	1.99	1.76
	Control group	0.27	(0.21)	0.33	(0.22)		0.06	0.75	0.28	0.56
5-back auditory	Experimental group	0.21	(0.13)	**0.49**	(0.14)	**,a	0.28	0.29	2.07	2.45
	Control group	0.28	(0.17)	0.2	(0.19)		0.08	0.52	0.44	0.64
0-back visuospatial	Experimental group	0.9	(0.09)	0.87	(0.11)		0.03	0.51	0.30	0.43
	Control group	0.97	(0.04)	**0.88**	(0.08)	*	0.09	-0.30	1.42	1.25
1-back visuospatial	Experimental group	0.82	(0.15)	0.76	(0.36)		0.06	0.08	0.22	0.23
	Control group	0.86	(0.1)	0.95	(0.04)		0.09	-0.23	1.18	1.06
2-back visuospatial	Experimental group	0.72	(0.15)	0.73	(0.13)		0.01	0.44	0.07	0.10
	Control group	0.7	(0.17)	0.75	(0.18)		0.05	0.10	0.29	0.30
3-back visuospatial	Experimental group	0.46	(0.19)	**0.85**	(0.09)	***,a	0.39	0.23	2.62	2.99
	Control group	0.46	(0.09)	0.37	(0.29)		0.09	0.43	0.42	0.55
4-back visuospatial	Experimental group	0.26	(0.15)	**0.73**	(0.12)	***,a	0.47	-0.37	3.92	3.35
	Control group	0.3	(0.18)	0.26	(0.24)		0.04	0.48	0.19	0.26
5-back visuospatial	Experimental group	0.2	(0.06)	**0.66**	(0.13)	***,a	0.46	-0.18	4.54	4.19
	Control group	0.22	(0.21)	0.31	(0.23)		0.09	0.58	0.41	0.63
Digit Span										
Forward	Experimental group	5.29	(0.76)	**6.00**	(0.82)	**	0.71	0.81	0.90	2.06
	Control group	5.88	(0.83)	6.38	(1.06)		0.5	0.55	0.53	0.78
Backwards	Experimental group	4.71	(0.95)	**6.29**	(1.38)	**	1.58	0.33	1.33	1.62
	Control group	6.00	(1.31)	5.13	(2.3)		0.87	0.71	0.46	0.86

Table 16. (Continued).

		Pre Test		Post Test			Absolute Mean Difference (pre vs. post)	Test-Retest Rel. (*r*)	*ES* (Cohen's d)	*ES* (corrected)
		M	**(SD)**	**M**	**(SD)**					
Vs2										
Forward	Experimental group	6.57	*(0.98)*	6.00	*(1)*		0.57	0.51	0.58	0.82
	Control group	6.38	*(1.41)*	6.88	*(1.13)*		0.5	0.03	0.39	0.40
Backwards	Experimental group	6.57	*(1.62)*	6.29	*(0.95)*		0.28	-0.02	0.21	0.21
	Control group	6.00	*(1.51)*	6.5	*(1.2)*		0.5	0.24	0.37	0.42
RST										
Span	Experimental group	2.36	*(0.56)*	**3.21**	*(0.64)*	**	0.85	0.57	1.41	2.16
	Control group	2.88	*(0.64)*	3.06	*(0.62)*		0.18	0.74	0.29	0.56
Task Switching (ms)										
Non-switch	Experimental group	1039	*(125)*	**886**	*(111)*	**	153	0.61	1.29	2.08
	Control group	1052	*(205)*	974	*(197)*		78	0.84	0.39	0.98
Switch	Experimental group	1309	*(214)*	**1058**	*(144)*	*	251	0.14	1.54	1.66
	Control group	1304	*(324)*	1213	*(385)*		91	0.62	0.26	0.41
Switch-Costs	Experimental group	271	*(206)*	172	*(131)*		99	-0.58	0.57	0.46
	Control group	252	*(137)*	239	*(209)*		13	0.18	0.07	0.08
Stroop (s)										
Reading	Experimental group	25.43	*(2.64)*	25.71	*(4.11)*		0.28	0.83	0.08	1.30
	Control group	25.38	*(2.83)*	24.38	*(3.07)*		1	0.46	0.34	0.03
Colour naming	Experimental group	39.43	*(3.69)*	**35.29**	*(3.82)*	***,a	4.14	0.86	1.15	0.22
	Control group	39.63	*(3.07)*	39.13	*(3.31)*		0.5	0.72	0.12	0.64
Interference	Experimental group	59.57	*(7.23)*	**55.57**	*(7.74)*	*	4	0.85	0.53	3.00
	Control group	60.5	*(6.09)*	60.63	*(7.05)*		0.13	0.62	0.02	0.19
Digit Symbol Test (Number of items)										
	Experimental group	81	*(8.93)*	80.57	*(10.28)*		0.43	0.78	0.04	0.09
	Control group	70.75	*(10.14)*	**78**	*(8.23)*	*	7.25	0.78	0.79	1.67
Priming										
General Speed (Picture naming)	Experimental group	5622	*(682)* a	5980	*(563)* a		358	0.38	0.57	0.73
	Control group	6838	*(1240)*	7069	*(798)*		231	0.44	0.22	0.30
Priming (%)	Experimental group	47.34	*(13.89)*	53.26	*(4.87)*		5.92	0.08	0.57	0.59
	Control group	50.27	*(15.32)*	55.7	*(7.33)*		5.43	0.77	0.45	0.95
Explicit Memory (Word list; Items recalled)										
Immediate recall	Experimental group	11.57	*(3.41)*	**14.43**	*(2.44)*	**	2.86	0.89	0.96	2.88
	Control group	13.25	*(4.59)*	13.5	*(3.34)*		0.25	0.76	0.06	0.13
Late Recall	Experimental group	9.86	*(3.44)*	**14**	*(2.31)*	***	4.14	0.78	1.41	2.99
	Control group	12	*(4.69)*	12	*(3.82)*		0	0.81	0.00	0.00
Intelligence Test (BOMAT)										
	Experimental group	10.14	*(2.67)*	**14.71**	*(3.55)*	***	4.57	0.83	1.45	3.55
	Control group	11.88	*(2.85)*	**13.88**	*(3.64)*	*	2	0.77	0.61	1.27

Note. RTs are reported as an average of *median* ms; Experimental group: *N* = 7; Control group: *N* = 8; *M* = Mean; *SD* = standard deviation; *ES* = Effect size; RST: Reading Span Task; Vs2: Visuospatial span. * *p* < 0.05; ** *p* ≤ 0.01; *** *p* ≤ 0.00: difference between pre and post session (one-tailed for the experimental group; two-tailed for the control group); significant differences between pre and post sessions are further indicated with bold font. *t*-values for the experimental group are between *t*(6) = 2.41 (*p* < 0.05) and *t*(6) = - 7.78 (*p* < 0.001) and for the control group between *t*(7) = -2.43 (*p* < 0.05) and *t*(7) = 3.97 (*p* < 0.01). a: Significant difference between experimental and control group (post session: One-tailed; pre session: Two-tailed); *t*-values are between *t*(13) = 2.79 (*p* < 0.01) and *t*(11.27) = 4.82 (*p* < 0.001) for the n-back tasks (P$_r$); and between *t*(13) = -2.20 (*p* < 0.05) and *t*(8.84) = -4.62 (*p* < 0.001) for the corresponding RTs. In the other measures, the t-values are between *t*(13) = -2.23 (*p* < 0.05) and *t*(13) = -3.01 (*p* < 0.001).

Effect sizes: Generally, the *ES* (Cohen's d) for the various measures are very large in the experimental group compared to those of the control group (see Table 16), providing even more evidence, that the n-back training had considerable effects on the

trained measure, with the highest *ES* being observed in the visuospatial 4- and 5-back tasks (accuracy) yielding values of 3.92 and 4.54 (compared to 0.19 and 0.41 in the control group), meaning that in the post session, more than 99.9% of the experimental group performs above the average obtained in the pre session (Bortz, 1999). Impressive *ES* were also observed on the various other WM tasks (digit-span test: *Forward*: 0.90 vs. 0.53; *backwards*: 1.33 vs. 0.46; RST: 1.41 vs. 0.29; always experimental vs. control group), but also extending into ability measures, such as explicit memory (*immediate recall*: 0.96 vs. 0.06; *late recall*: 1.41 vs. 0.00; experimental vs. control group) and intelligence. In the intelligence measure, although that an effect was observed in both groups, the *ES* was more than twice as large in the experimental group than in the control group (1.45 vs. 0.61).

Reliability measures: The test-retest reliability in the n-back considerably varied between groups (see Table 16) and generally yielded larger values in the RTs than in the accuracies. To some extent, this low reliability in these tasks most probably results from the restricted variance due to ceiling and floor levels at higher and lowest levels of load, but probably also, due to the small sample size (see Experiment 1). Similarly, in the Vs2 and in the task switching paradigm, very low and inconsistent reliability values were observed, which might have restricted the possibility to find a significant treatment effect in these measures. In the priming paradigm, the experimental group showed near-zero reliabilities, as is consistent with findings by Meier (1999) or Meier and Perrig (2000), whereas the control group showed very high reliability scores in this task. In the other tasks, the reliability values are quite consistent between the two groups.

7.3.4. Discussion

Adaptive training of a dual n-back task yielded significant improvements not only in the trained task, but also in a wide range of cognitive ability, WM tasks and speed; a finding which has not been reported to this extent before. The findings in explicit memory for example are striking, and to my knowledge, there is no study reporting such effects following a WM or attention training. The *g*-improvements are consistent with data reported by the Klingberg-group (Klingberg et al., 2005; Klingberg et al., 2002; Olesen et al., 2004) and by Posner and Rothbart (2005). However, as suspected, an improvement has also been obtained in the control group, suggesting that in such a task, re-testing might pose a problem, and is also commonly observed in other 'executive' tasks (e.g. Rabbitt, 1997). Still, the significant interaction between session and group, as well as the larger *ES* in the experimental group indicate that some variance can be clearly attributed to the n-back training. It is also of importance, that the n-back training has differential effects, i.e., as predicted, no treatment-effects were observed in the digit-symbol test and in the used priming task. Conversely, the control group performed significantly better in the digit-symbol test in the post session, but this can be explained with the very low performance in the pre session.

In contrast to our hypothesis, no effects were observed in the Vs2, which can be explained theoretically and methodologically: Although developed in order to assess visuospatial WM, there might be other processes than WM processes involved in this task. As for example Schellig (1993) states, this task also is derived for the assessment of the capacity of the visuospatial component of WM *and* of implicit visual-spatial learning. There is evidence that visual material automatically engages multiple representations and associations, thus relying on more elaborate and variable processes in encoding than verbal material (e.g. Grady, McIntosh, Rajah, & Craik, 1998). These multiple processes coming into play in such a task might already be optimal; therefore, the WM training did not affect the performance in the Vs2. On methodological basis, there is clearly a problem of reliability, which most probably resulted through the randomly presented locations: In some trials, the track to follow might be very easy, while in others, it might be more challenging. With such a small sample therefore, this procedure has not been optimal and would have to be optimized for a future study.

Similarly, the absent effects in the switch costs despite the much faster RTs in the post session in the experimental group might either be attributable to different processes involved in this task than 'executive' processes (e.g. Meiran et al., 2000), or to the low reliability of the task. However, the larger *ES* in the experimental group suggest that a large part of the variance can be explained through the WM training.

In sum, these findings have an important impact on the discussion regarding the concept of *g* and other basic cognitive ability measures, such as memory and attention: With this data, the assumption of a fixed capacity is clearly challenged. The results cannot be explained solely with applied strategies, since no strategies were taught to the participants. It is well possible, that strategies came into play during the training sessions of the n-back task, but these strategies are certainly not strategies based on representations in long-term memory, as observed for example in experts during the digit-span task (Ericsson & Chase, 1982): In the n-back task, continuous updating is required; every three seconds two new randomly determined stimuli have to be memorized, and at the same time, the two stimuli presented *n* +1 positions before, have to be discarded. Therefore, the sequence to be memorized changes every three seconds and is not static as in the example of the digit-span task. If strategies based on long-term representations would have been applied, these representations should be recalled afterwards (Ericsson & Delaney, 1998). Although we did not test this explicitly, it seems to be very unlikely, since the interference between new to-be-remembered sequences would be enormous. Moreover, as mentioned before, the acquired expertise of the participants described by Ericsson & Chase (1982) did not extend into another domain, therefore remaining task specific and speaking for different processes being involved with such a training as conducted here.

Further evidence that not the use of strategies, but enhanced WMC was responsible for improvement in the cognitive ability tasks is provided by the control group: If strategies were applied systematically in the post test by the participants, some improvements would have also been observed in the control group (e.g. in explicit memory), moreover since the participants knew that they would have to perform similar tasks in the post session.

The question remains how it is possible to be able to reach such a level of proficiency in the dual n-back task in such a short time, when most people approach a floor level already at the 3-back condition. What mechanisms might underlie such a

performance? Frequency judgements in the sense of Hasher & Zacks (1979) might be considered as possible processes coming into play in an n-back task, since better performance is obtained if the task is executed 'automatic' and 'effortless', which is also confirmed by the participants' reports (see Experiment 2): The possible targets have to be recognized 'intuitively' as being the same as n positions back in the sequence; especially at high levels of load, active rehearsal does not work anymore. However, there is evidence against similar processing in frequency judgements and the dual n-back task: In frequency judgements, usually no age effects are observed (Hasher & Chromiak, 1977), which is not in accordance with results of Experiment 4. Also, according to Hasher & Chromiak (1977), there is no indication, that frequency judgements can be improved with training and no relationship between intelligence, i.e., ability differences are observed (Zacks, Hasher, & Sanft, 1982). More probable explanations are provided by Duncan (1980), who mentions strategies like grouping or Gestalt-like processes which can be used to extend capacity limitations. Such processes are in accordance with the reports of the participants: Although they find it very difficult to verbalize their strategies; the most common described approach was to somehow 'bind' the verbal and the visuospatial stimuli together to one stimulus in order to save resources.

To conclude, our data provide much evidence that basic cognitive abilities can benefit from adaptive n-back dual-task training, either through an increase of WMC or through an increase in processing efficiency. These findings open a wide range in practical application domains, such as retraining after brain lesions, training children with ADHD, or training people with learning disabilities. Still, it remains to be seen, whether the obtained training effects remain over a longer, training-free time, therefore a follow-up study will shed more light on that matter.

8. General Discussion and Conclusion

In this thesis I have been looking upon behavioural and neuronal correlates of capacity limitations in human cognition. I have shown that the n-back task proves to be a valid and reliable instrument in order to investigate these questions in yielding large and replicable interindividual differences in performance, especially in the dual-task versions at high levels of load. These individual differences were also predictive for other higher-order cognitive tasks, i.e. for performance in Raven's APM. By means of fMRI, I demonstrated that load-dependent processing is reflected on many areas in the brain, but also, that the prefrontal cortices seem to be especially affected when capacity limitations are reached. However, prefrontal areas did not only reflect load-dependent processing, but also interindividual differences at the edge of capacity: While some individuals relied on the resources of the PFC to a very large extent when performing a complex task, there were others who processed very efficiently, as was expressed in very low activation patterns in most cortical areas. It was hypothesized that these differences in activation might be related to the dissociation between controlled and automatic processing: While some participants direct a huge amount of attentional resources to these tasks, others seem to perform these tasks in a more automatic, but nevertheless very efficient way. In advancing age, the reliance on prefrontal areas seems even more important, being reflected in neuronal compensatory mechanisms in addition to the prevalent controlled processing in order to fulfil task demands. Despite demonstrating these interindividual differences, my data do not provide enough information in order to clarify, why there are some participants who are better in those tasks, and why there are others, who fail. One clue might be found in general intelligence, being correlated fairly well with the performance in the n-back task. However, the shared variance is too small in order to provide a full explanation. Where might be the other sources of variance? Ignoring claims of biologically and genetically determined constraints in capacity, I decided to investigate, whether it would be possible to 'produce' high-performing participants by means of very demanding dual-task training. The results of these experiments were striking: Average-performing participants reached an impressive proficiency in this task after just a couple of training days. Yet, even more impressing were the observed generalizing effects onto other ability measures, which have not been shown to this extent before. These generalizing effects might reflect neuronal plasticity in

association areas, which would be very challenging to tackle by means of fMRI or EEG (see below). Thus, the potential of such a training seems respectable and does also open a wide range of practical applications, for example in the domain of children with learning disabilities, where plasticity is likely to be even more pronounced than in adults (e.g. Qin et al., 2004).

Thus, returning to my initial question: Are there capacity limitations in human cognition? In looking at the training data it is tempting to assume that there are no capacity limitations. Still, also in these participants, the underlying processes yielding such proficiency are not clear, and I can only speculate that the training might facilitate efficiency in processing, resulting in fast formation of associations, which are important in those tasks being applied in the post session and being expressed in the faster RTs. It is also not evident, how such a training might aid to coping with everyday life, an issue which is very hard to tackle as being pointed out by Sturm et al. (2003). Nevertheless, with such training there are clearly processes initiated, which seem to enhance efficacy which in turn might facilitate a wide range of cognitive processes.

On the other hand, one might be ask, why would it be desirable to extend capacity limitations? Or is there a reason, why there are capacity limitations in the first place? Why it would be necessary to place constraints on a system? In asking these questions, I have come across a very different view of capacity limitations as was discussed until now and I will briefly provide some reflections to the topic of functional significance of capacity limitations: On biological basis, there are several conditions providing information processing capacity. First, there is the large number of neurons provided in the human brain, consisting of approximately one hundred billions (1.2×10^{10}) of neurons and exponentially more connections, being reflected in the amount of 3.6×10^{14} synapses (Marois & Ivanoff, 2005; Roth & Dicke, 2005). These neurons and their synaptic connections, which are not very distant in the human brain, provide a massive potential of parallel processing. But the number of neurons is not the only determinant in providing processing capacity: There is also the conduction velocity of cortical fibres, mostly determined by the diameter of myelinated fibres, which seems crucial for fast and efficient processing. Compared to non-primates, these myelinated cortical fibres are particularly thick in humans and other primates (Roth & Dicke, 2005). Therefore, the amount of myelinated fibres,

combined with the cell density, and with the relatively small distances between neurons, the best environment for almost unlimited capacity in information processing is provided and is as such uniquely represented in the human brain. A capacity limit in human information processing is therefore not necessarily assumed at first thought; although, as I have shown, there are many capacity limitations, which are quite obvious in our everyday life, already when we fail to perform two relatively simple tasks simultaneously. So, where does this capacity limitation come from? Are there any advantages in posing constraints on a system?

To my view, the advantage of constraints can be explained as follows: On one hand, a most natural physical and evolutionary constraint is placed on our brain in terms of relative size: The human brain accounts only for 2% of the body mass, but nevertheless consumes 20% of the total metabolism, which has to be maintained with energy, i.e. with food (Roth & Dicke, 2005). Therefore, our brain is very 'expensive' in terms of energy supply. Increasing capacity in increasing brain size therefore does not seem to be an evolutionary advantage, since in increasing brain size, also the intake of food has to be considerably modified, and further, the balance between the demands of other organs has to be maintained.

There is evidence, that evolution has already encountered such a problem, since the human brain size has linearly increased in relation to body size during the 3.5 million years of human evolution, from a volume of 600 cm^3 in the Homo habilis, to a volume of 1500 cm^3 in the extinct Homo neanderthalensis. However, this brain size did not seem to facilitate survival, since interestingly, the brain of the Homo neanterthalensis was considerably larger than that of the modern Homo sapiens with its 1350 cm^3 (Roth & Dicke, 2005). It seems therefore that the evolution has produced the optimal solution as to the proportion of brain and body size in the Homo sapiens, thus suggesting that the constraints in human brain size might have its particular reason. One might therefore consider the following explanations: Limitations may force an organism to selectivity, that is, the available hardware (e.g. neurons and their interconnections) has to be used in the most efficient way and therefore, only the most relevant information is to be processed. In placing constraints to the amount of information being processed, the organism has to select and integrate only the optimal solution or stimulus, in order to assure the survival and the fitness in regard to other organisms. This selectivity has to be fast (e.g. a predator has to be recognized in a

fraction of a second; or more psychologically, an emotion of a counterpart has to be processed instantaneously), and therefore, it is rather uneconomic to process unnecessary other inputs, such as the colour of its eyes, or to guide the attention to the sunset behind it; all we have to do is to discriminate, recognize and classify, in order to select the correct and adequate consequences (either flee or fight, or mirror the counterparts' emotion). Learning to discriminate and to classify in order to make sense of our environment is one of the first things we have to learn in our life, and this discrimination and selectivity is also reflected in the development of the brain: While there is an enormous and rapid increase in cells in early development (proliferation), almost half the neurons generated in the mammalian nervous system are lost through a process known as apoptosis, i.e. programmed cell death (e.g. Jessell & Sanes, 2000). The changes and the rates of cell proliferation and/or apoptosis initiate the development and differentiation of the cerebral cortex in the formation of radial columns (Rakic, 2005), followed by areal specification (Rakic, 1988, 2002). These processes are guided by complex genetic and molecular mechanisms, which go beyond the scope of this thesis and I therefore refer to Kandel, Schwartz, and Jessell (2000).

Now, how might these morphological differentiation processes are reflected physiologically and psychologically? It is widely acknowledged that one of the central problems in cognitive neuroscience is to determine how cognitive processes, and in this context, how capacity limitations in information processing are derived from brain processes. In physiological terms, it is the synchronized firing of cell assemblies which reveals much about the nature of cognitive processes (e.g. Ward, 2003) and seems also providing the basis for selective and optimal task performance. The formation of cell-assemblies is either evolutionary defined or acquired through experience in order to adapt the system most optimally to the requirements of the environment (Edelman & Tononi, 1998). Specifically, as larger the amount of components (e.g. individual cells, or cell assemblies), and as more intensive and non-linear the interaction between them, the more emphasis has to be placed on selective mechanisms in order to be adaptive (Edelman & Tononi, 1998).

In psychological and less evolutionary terms, the benefit of selectivity can be also illustrated with heuristics applied in problem-solving (Groner, Groner, & Bischof, 1983; Newell, 1983): With the use of heuristics, the so called problem space

can be markedly reduced by taking into account only the relevant information, and can therefore lead to the solution very fast and efficiently (however, with heuristics, the correct solution is not always obtained; depending on the knowledge of the problem-solver). Conversely, with the use of algorithms, the system is depending on a well defined sequence of operation criteria, including the complete problem space and generating a solution with certainty for the defined problem class, which in turn can be very time-consuming.

Another related explanation for the necessity of selectivity evolves from a historical case study by Luria (1968): One of his patients, 'S.' was not able to forget, rather, every piece of information was stored in his obviously enormous memory. What would be considered as practical at first thought was very limiting for S., and he was not able to live an independent life anymore: He could not differentiate between important and less important things, an ability which seems also to be relevant in social interaction. This is in line with descriptions of patients with frontal damage, which also sometimes have troubles with selective attention, i.e., in discriminating between relevant topics in discourse and become very associative and even logorrhoic in some cases (Fuster, 1989). Another interesting phenomenon of S.'s prodigious memory, which seems also related to the advantage of capacity constraints, was that S. was virtually paralyzed when it came to understanding poetry, since metaphorical thinking was almost impossible for him: He lived in a world of unique particulars and was therefore not able to think in abstract terms. As Luria wrote, "S. found that when he tried to read poetry the obstacles to his understanding were overwhelming: Each expression gave rise to an image; this, in turn, would conflict with another image that had been evoked." (Luria, 1968; p. 120).

To conclude, the optimal balance between remembering and forgetting gives support to the selectivity of the system in order to be able to focus on relevant inputs (and outputs). Thus, capacity constraints come to aid to the economics of the human information processing system and therefore seem to be highly relevant for our ability to cope with life. It seems that the better performance in high-performing or trained participants does not evolve from a larger capacity, but rather, from the more efficient and selective processing, being possibly reflected in more synchronous neural oscillations of nerve cells (see below).

Table 17. Summary of capacity limitations and their sources and challenges; adapted and extended from Kintsch, Healy, Hegarty, Pennington, and Salthouse (1999).

Sources for Constraints	Authors
***Structural** or task-related constraints*	
(Time-based) Information decay	Baddeley and Logie (1999); Kieras, Meyer, Mueller, and Seymour (1999)
Modality-specific interference	Wickens (1980; 1984)
Extensive perceptual analysis	Pashler (1998)
Limitations in communications and interactions among different subsystems or subcomponents	Schneider (1999)
Limits in the availability of activation	Engle et al. (1999); Lovett, Reder, and Lebiere (1999)
(Similarity-based) Interference	Baddeley and Logie (1999); O'Reilly, Braver, and Cohen (1999)
Psychological Refractory Period	Telford (1931); Pashler (1998)
*Constraints due to **interindividual** variability*	
Efficiency of controlled attention and/or executive mechanisms	Baddeley and Logie (1999); Engle et al. (1999)
Limits in processing speed or efficiency	Salthouse (1996); Salthouse, Atkinson, and Berish (2003)
Lack of inhibitory control	Engle et al. (1999); O'Reilly et al. (1999); Kane, Hasher, Stoltzfus, Zacks, and Connelly (1994); Zacks and Hasher (1994)
Lack of skill or knowledge for efficient encoding and retrieval	Ericsson and Kintsch (1995); Ericsson and Delaney (1998); Engle et al. (1999)
Lack of intelligence	Engle et al. (1999)
***Biological** constraints*	
Genetic factors	Baddeley and Logie (1999); Ericsson and Lehmann, 1996; O'Reilly et al. (1999); Goldberg and Weinberger (2004)
Morphological factors	Roth and Dicke (2005); Rakic (2005)
Development	Baddeley and Logie (1999); Pascual-Leone (1979); Pascual-Leone and Johnson (2005)
Neural functioning or efficiency (e.g. PFC)	Callicott et al. (1999); Engle et al. (1999); Friedman, Campbell Polson, Dafoe, and Gaskill (1982)
Acquired pathologies (e.g. brain damage, schizophrenia)	Baddeley and Logie (1999)
Neuropharmacological factors (e.g. dopamine)	Egan et al. (2001); Goldberg and Weinberger (2004)
***Other** (unspecific) constraints*	
Motivation	Duncan (1980) ; Pochon et al. (2002) ; Heitz et al. (2005), Visser and Merikle (1999)
Effort	Kahneman (1973)
Arousal	Yerkes and Dodson (1908) ; Kahneman (1973); Matthews and Deary (1998)
Emotion	Peretti (1998)
Stress	Hockey (1984); Beilock and Carr (2005)
Personality traits	Eysenck (1982); Humphreys and Revelle (1984); Matthews and Deary (1998); Posner and Rothbart (2005); Gignac (2005); DeYoung, Peterson, and Higgins (2005)

In this work, I have shown that capacity limitations are dependent on interindividual differences, which in turn are shaped by development, learning, strategies, intelligence, but also by brain function, genetics and evolution. In fact, there are many sources of capacity constraints, and it was not possible to discuss all of them to a full extent. But I will nevertheless provide a summary of capacity constraints brought up in the literature as shown in Table 17, going beyond the issues discussed here, but which is still most certainly incomplete. This summary might also illustrate that there are many issues to be solved, thus making it clear that investigating and explaining the mechanisms underlying capacity limitations remain a challenge to researchers, also bearing in mind that capacity limitations might not be disadvantageous.

Finally, what might be the prospects for future research? Of course, there are still many issues to be addressed: For instance, the impact of emotion and/or personality traits on load-dependent processing and on the related capacity limitations could be investigated. In the literature, there is indeed some evidence for differential effects in WMC which can be attributed to personality traits (see Matthews & Deary (1998) for a review, or DeYoung, Peterson & Higgins (2005) for a more recent finding). The interaction of personality traits and WMC might also be reflected on neuronal level, as has been shown with EEG-data (Matthews & Amelang, 1993), but which might also be traceable with fMRI. The potential of EEG could also be taken advantage of, especially regarding the question of processing efficiency as expressed in the high-performing and trained individuals. Since in the fMRI-data, no consistent brain area has been found which could be interpreted as compensating for the non-existent prefrontal activation in high-performing individuals, it could be hypothesized that this efficiency might be tackled with spectral analyses by means of EEG. Especially gamma band activity might be a likely candidate in order to clarify this question, since it is assumed to reflect attentional processes, and further, that synchronous neural oscillations at the gamma frequency is seen as being the neural correlate of conscious awareness (see Ward, 2003 for a review). As stated before, I would hypothesize that high-performing and trained participants would show markedly more synchronous firing at the gamma frequency than low-performing participants, also taking into account the reported 'intuitive' strategy used in this group.

Thus, there are many open questions and many issues to be solved, however I will stop here to conclude with Einstein, in the International Year of Relativity:

The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvellous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity.

Albert Einstein

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin, 131*(1), 30-60.

Adcock, R. A., Constable, R. T., Gore, J. C., & Goldman-Rakic, P. S. (2000). Functional neuroanatomy of executive processes involved in dual-task performance. *Proceedings of the National Academy of Sciences of the United States of America, 97*(7), 3567-3572.

Allport, A. D., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*, 225-235.

Allport, D. A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science*. Cambridge, MA: MIT Press.

Anourova, I., Rama, P., Alho, K., Koivusalo, S., Kahnari, J., & Carlson, S. (1999). Selective interference reveals dissociation between auditory memory for location and pitch. *Neuroreport, 10*(17), 3543-3547.

Attneave, F., & Arnoult, M. D. (1956). Methodological considerations in the quantitative study of shape and pattern perception. *Psychological Bulletin, 53*, 221-227.

Baddeley, A. D. (1986). *Working Memory*. Oxford, England: Clarendon Press.

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology, 49A*(1), 5-28.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-89): Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working Memory: The Multiple-Component Model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 28-61). Cambridge: Cambridge University Press.

Barch, D. M., Sheline, Y. I., Csernansky, J. G., & Snyder, A. Z. (2003). Working memory and prefrontal cortex dysfunction: specificity to schizophrenia compared with major depression. *Biological Psychiatry, 53*(5), 376-384.

Bark, N., Revheim, N., Huq, F., Khalderov, V., Ganz, Z. W., & Medalia, A. (2003). The impact of cognitive remediation on psychiatric symptoms of schizophrenia. *Schizophrenia Research, 63*(3), 229-235.

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., et al. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10*(2), 344-380.

Bäumler, G. (1985). *Farbe-Wort-Interferenztest (FWIT) nach J.R. Stroop*. Göttingen: Hogrefe.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: working memory and "choking under pressure" in math. *Psychological Science, 16*(2), 101-105.

Bell, M., Bryson, G., Greig, T., Corcoran, C., & Wexler, B. E. (2001). Neurocognitive enhancement therapy with work therapy: effects on neuropsychological test performance. *Archives of General Psychiatry, 58*(8), 763-768.

Bellack, A. S., Gold, J. M., & Buchanan, R. W. (1999). Cognitive rehabilitation for schizophrenia: problems, prospects, and strategies. *Schizophrenia Bulletin, 25*(2), 257-274.

Bellack, A. S., Weinhardt, L. S., Gold, J. M., & Gearon, J. S. (2001). Generalization of training effects in schizophrenia. *Schizophrenia Research, 48*(2-3), 255-262.

Ben-Yishay, Y., Piasetzky, E. B., & Rattok, J. (1987). A systematic method for ameliorating disorders in basic attention. In M. J. Meier & A. L. Benton & L. Diller (Eds.), *Neuropsychological Rehabilitation*. Edinburgh: Churchill Livingstone.

Binet, A. (1894). *Psychologie des Grand Calculateurs et Joueurs d'Echecs*. Paris: Librairies Hachette.

Binet, A. (1903). *L'étude expérimentale de l'intelligence*. Paris: Schleicher Frères.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5th ed.). Berlin: Springer.

Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 49A*, 525–545.

Braff, D. L. (1993). Information processing and attention dysfunctions in schizophrenia. *Schizophrenia Bulletin, 19*(2), 233-259.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 5*(1), 49-62.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review, 64*, 205-214.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in Long-Term Memory* (pp. 3-18). New York: Wiley.

Bunge, S. A., Klingberg, T., Jacobsen, R. B., & Gabrieli, J. D. (2000). A resource model of the neural basis of executive working memory. *Proceedings of the National Academy of Sciences of the United States of America, 97*(7), 3573-3578.

Burgess, P. W. (1997). Theory and Methodology in Executive Function Research. In P. Rabbitt (Ed.), *Methodology of Frontal and Executive Function* (pp. 81-116). Hove: Psychology Press.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and Aging, 17*, 85-100.

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.

Cabeza, R., & Nyberg, L. (2003). Seeing the forest through the trees: The cross-function approach to functional neuroimaging. In A. Zani & A. M. Proverbio (Eds.), *The Cognitive Electrophysiology of Mind and Brain*. San Diego: Academic Press.

Calev, A. (1984). Recall and recognition in chronic nondemented schizophrenics: Use of matched tasks. *Journal of Abnormal Psychology, 93*(2), 172-177.

Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., et al. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex, 9*(1), 20-26.

Carlson, S., Martinkauppi, S., Rama, P., Salli, E., Korvenoja, A., & Aronen, H. J. (1998). Distribution of cortical activation during visuospatial n-back tasks as

revealed by functional magnetic resonance imaging. *Cerebral Cortex, 8*(8), 743-752.

Carpenter, P. A., Just, M. A., Keller, T. A., Eddy, W., & Thulborn, K. (1999). Graded functional activation in the visuospatial system with the amount of task demand. *Journal of Cognitive Neuroscience, 11*(1), 9-24.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153-193.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 215-281). New York: Academic.

Cherry, C. E. (1953). Some experiments on the reception of speech with one and two ears. *Journal of the Acoustical Society of America, 25*, 975-979.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collette, F., Olivier, L., Van der Linden, M., Laureys, S., Delfiore, G., Luxen, A., et al. (2005). Involvement of both prefrontal and inferior parietal cortex in dual-task performance. *Cognitive Brain Research, 24*, 237– 251.

Conway, A. R. (2005). Cognitive mechanisms underlying intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 47-59). Thousand Oaks: Sage Publications.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychon Bull Rev, 8*(2), 331-335.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114; discussion 114-185.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42-100.

D'Esposito, M. (2001). Working Memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional Neuroimaging of Cognition* (pp. 293-327). Cambridge, Massachusetts: MIT Press.

D'Esposito, M., Detre, J. A., Alsop, D. C., Shin, R. K., Atlas, D., & Grossmann, M. (1995). The neural basis of the central executive system of working memory. *Nature, 378*, 279-281.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 561-584.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychon Bull Rev, 3*(4), 422-433.

de Groot, A. D. (1946). *Thought and Choice and Chess*. Den Haag (NL): Mouton.

Della-Maggiore, V., Grady, C. L., & McIntosh, A. R. (2002). Dissecting the effect of aging on the neural substrates of memory: deterioration, preservation or functional reorganization? *Reviews in the Neurosciences, 13*, 167-181.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2005). Sources of openness/intellect: Cognitive and neuropsychological correlates of the fifth factor of personality. *Journal of Personality, 73*(4), 825-858.

Druzgal, T. J., & D'Esposito, M. (2001). Activity in fusiform face area modulated as a function of working memory load. *Brain Research. Cognitive Brain Research, 10*(3), 355-364.

Duncan, J. (1980). The demonstration of capacity limitation. *Cognitive Psychology, 12*, 75-96.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*(2), 170-177.

Edelman, G. M., & Tononi, G. (1998). Neuronaler Darwinismus: Eine selektionistische Betrachtungsweise des Gehirns. In H. Meier & D. Ploog (Eds.), *Der Mensch und sein Gehirn. Die Folgen der Evolution* (pp. 187-233). München: Piper.

Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., et al. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America, 98*(12), 6917-6922.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 102-134). Cambridge: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Ericsson, A. K., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review, 102*(2), 211-245.

Ericsson, K. A. (1988). Analysis of memory performance in terms of memory skill. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 137-179). Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. *American Scientist, 70*(6), 607-615.

Ericsson, K. A., & Delaney, P. F. (1998). Working memory and expert performance. In R. H. Logie & K. J. Gilhooly (Eds.), *Working Memory and Thinking* (pp. 93-114). Hillsdale, NJ: Erlbaum.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.

Eysenck, M. W. (1982). *Attention and Arousal: Cognition and Performance*. Berlin: Springer.

Fisk, J. E., & Warr, P. (1996). Age and working memory: the role of perceptual speed, the central executive, and the phonological loop. *Psychology and Aging, 11*(2), 316-323.

Fiszdon, J. M., Bryson, G. J., Wexler, B. E., & Bell, M. D. (2004). Durability of cognitive remediation training in schizophrenia: performance on two memory tasks at 6-month and 12-month follow-up. *Psychiatry Research, 125*(1), 1-7.

Fletcher, P. C., Frith, C. D., Grasby, P. M., Shallice, T., Frackowiak, R. S., & Dolan, R. J. (1995). Brain systems for encoding and retrieval of auditory-verbal memory. An in vivo study in humans. *Brain, 118*, 401-416.

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

Fogarty, G., & Stankov, L. (1982). Competing tasks as an index of intelligence. *Personality and Individual Differences, 3*, 407-422.

Friedman, A., Campbell Polson, M., Dafoe, C. G., & Gaskill, S. J. (1982). Dividing attention within and between hemispheres: testing a multiple resources approach to limited-capacity information processing. *Journal of Experimental Psychology: Human Perception and Performance, 8*(5), 625-650.

Friston, K. J., Mechelli, A., Turner, R., & Price, C. J. (2000). Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage, 12*(4), 466-477.

Fuster, J. M. (1989). *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe*. New York: Raven Press.

Gazzaley, A. H., & D'Esposito, M. (2005). BOLD functional MRI and cognitive aging. In R. Cabeza & L. Nyberg & D. Park (Eds.), *Cognitive Neuroscience of Aging. Linking Cognitive and Cerebral Aging* (pp. 107-131). Oxford: Oxford University Press.

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex, 10*, 829-839.

Gevins, A. S., Bressler, S. L., Cutillo, B. A., Illes, J., Miller, J. C., Stern, J., et al. (1990). Effects of prolonged mental work on functional brain topography. *Electroencephalography and Clinical Neurophysiology, 76*(4), 339-350.

Gignac, G. E. (2005). Openness to experience, general intelligence and crystallized intelligence: A methodological extension. *Intelligence, 33*, 161–167.

Goldberg, T. E., Berman, K. F., Fleming, K., Ostrem, J., Van Horn, J. D., Esposito, G., et al. (1998). Uncoupling Cognitive Workload and Prefrontal Cortical Physiology: A PET rCBF Study. *Neuroimage, 7*(4), 296-303.

Goldberg, T. E., & Weinberger, D. R. (2004). Genes and the parsing of cognitive processes. *Trends Cogn Sci, 8*(7), 325-335.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351*(1346), 1445-1453.

Grady, C. L., McIntosh, A. R., & Craik, F. I. (2005). Task-related activity in prefrontal cortex and its relation to recognition memory performance in young and old adults. *Neuropsychologia, 43*(10), 1466-1481.

Grady, C. L., McIntosh, A. R., Rajah, M. N., & Craik, F. I. (1998). Neural correlates of the episodic encoding of pictures and words. *Proceedings of the National Academy of Sciences of the United States of America, 95*(5), 2703-2708.

Gray, J. A., & Wedderburn, A. A. (1960). Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 12*, 180-184.

Gray, J. M., & Robertson, I. (1989). Remediation of attentional difficulties following brain injury: three experimental single case studies. *Brain Injury, 3*, 163-170.

Gray, J. M., Robertson, I., Pentland, B., & Anderson, S. (1992). Microcomputer-based attentional retraining after brain damage: a randomized group controlled trial. *Neuropsychological Rehabilitation, 2*, 97-115.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6*(3), 316-322.

Green, M. F. (1998). *Schizophrenia from a neurocognitive perspective*. Boston: Allyn and Bacon.

Groner, M., Groner, R., & Bischof, W. (1983). Approaches to heuristics: A historical review. In R. Groner & M. Groner & W. Bischof (Eds.), *Methods of Heuristics* (pp. 2-18). Hillsdale, NJ: Lawrence Erlbaum.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(6), 803-831; discussion 831-864.

Handy, T. C. (2000). Capacity theory as a model of cortical behavior. *Journal of Cognitive Neuroscience, 12*(6), 1066-1069.

Hasher, L., & Chromiak, W. (1977). The processing of frequency information: An automatic mechanism? *Journal of Verbal Learning and Verbal Behaviour, 16*, 173-184.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356-388.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *American Psychologist, 39*(12), 1372-1388.

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 61-77). Thousand Oaks, CA: Sage Publications.

Hockey, A., & Geffen, G. (2004). The concurrent validity and test-retest reliability of a visuospatial working memory task. *Intelligence, 32*, 591-605.

Hockey, G. R. J. (1984). Varieties of attentional state: the effects of the environment. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention*. New York: Academic Press.

Hoff, A. L., Sakuma, M., Wieneke, M., Horon, R., Kushner, M., & DeLisi, L. E. (1999). Longitudinal neuropsychological follow-up study of patients with first-episode schizophrenia. *American Journal of Psychiatry, 156*(9), 1336-1341.

Hoshi, Y., Tsou, B. H., Billock, V. A., Tanosaki, M., Iguchi, Y., Shimada, M., et al. (2003). Spatiotemporal characteristics of hemodynamic changes in the human lateral prefrontal cortex during working memory tasks. *Neuroimage, 20*(3), 1493-1504.

Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizentest. BOMAT - advanced - short version*. Göttingen: Hogrefe.

Hossiep, R., Turck, D., & Hasella, M. (2001). *Bochumer Matrizentest. BOMAT - advanced - short version*.

Humphreys, M. S., & Revelle, W. (1984). Personality, motivation and performance: a theory of the relationship between individual differences and information processing. *Psychological Review, 91*, 153-184.

Jaeggi, S. M., Meier, B., & Buschkuehl, M. (in prep.). The N-back Task as Single and Dual-Task Version Associated with Working Memory, Prospective Memory, and Intelligence Measures: A Correlational Study (Working title).

Jaeggi, S. M., Seewer, R., Nirkko, A. C., Eckstein, D., Schroth, G., Groner, R., et al. (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: functional magnetic resonance imaging study. *Neuroimage, 19*(2), 210-225.

James, W. (1890). *The Principles of Psychology* (Vol. 1). New York: Holt.

Jansma, J. M., Ramsey, N. F., Coppola, R., & Kahn, R. S. (2000). Specific versus nonspecific brain activity in a parametric N-back task. *Neuroimage, 12*(6), 688-697.

Jansma, J. M., Ramsey, N. F., Slagter, H. A., & Kahn, R. S. (2001). Functional anatomical correlates of controlled and automatic processing. *Journal of Cognitive Neuroscience, 13*(6), 730-743.

Jensen, A. R. (1993). Why is reaction time correlated with psychometric g? *Current Directions in Psychological Science, 2*, 53-56.

Jessell, T. M., & Sanes, J. R. (2000). The generation and survival of nerve cells. In E. R. Kandel & J. H. Schwartz & T. M. Jessell (Eds.), *Principles of Neural Science* (4th ed., pp. 1041-1062). New York: McGraw-Hill.

Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, UK: Cambridge University Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General, 107*, 420-435.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., et al. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9*(4), 462-475.

Joy, S., Fein, D., Kaplan, E., & Freedman, M. (2000). Speed and memory in WAIS-R NI Digit Symbol performance among healthy older adults. *Journal of the International Neuropsychological Society, 6*, 770-780.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension. *Psychological Review, 99*, 122-149.

Just, M. A., Carpenter, P. A., Keller, T. A., Emery, L., Zajac, H., & Thulborn, K. R. (2001). Interdependence of Nonoverlapping Cortical Systems in Dual Cognitive Tasks. *Neuroimage, 14*(2), 417-426.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, New Jersey: Prentice-Halls.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (Eds.). (2000). *Principles of neural science* (4th ed.). New York: McGraw-Hill.

Kane, M. J. (2005). Full frontal fluidity? Looking in on the neuroimaging of reasoning and intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 141-163). Thousand Oaks: Sage Publications.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: limits on long-term memory retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*(2), 336-358.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon Bull Rev, 9*(4), 637-671.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackermann, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66-71.

Kane, M. J., Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Connelly, S. L. (1994). Inhibitory attentional mechanisms and aging. *Psychol Aging, 9*(1), 103-112.

Kassubek, J., Schmidtke, K., Kimmig, H., Lucking, C. H., & Greenlee, M. W. (2001). Changes in cortical activation during mirror reading before and after training: an fMRI study of procedural learning. *Brain Research. Cognitive Brain Research, 10*(3), 207-217.

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control* (pp. 183-223). Cambridge, UK: Cambridge University Press.

Kintsch, W., Healy, A. F., Hegarty, M., Pennington, B. F., & Salthouse, T. A. (1999). Models of working memory. Eight questions and some general issues. In A. Miyake & P. Shah (Eds.), *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control* (pp. 412-441). Cambridge, UK: Cambridge University Press.

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

Klingberg, T. (1998). Concurrent performance of two working memory tasks: potential mechanisms of interference. *Cerebral Cortex, 8*(7), 593-601.

Klingberg, T. (2000). Limitations in information processing in the human brain: neuroimaging of dual task performance and working memory tasks. *Progress in Brain Research, 126*, 95-102.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (2005). Computerized training of working memory in children with ADHD--a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(2), 177-186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24*(6), 781-791.

Klingberg, T., & Roland, P. E. (1997). Interference between two concurrent tasks is associated with activation of overlapping fields in the cortex. *Brain Research. Cognitive Brain Research, 6*(1), 1-8.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature, 399*(6732), 148-151.

Kondo, H., Osaka, N., & Osaka, M. (2004). Cooperation of the anterior cingulate cortex and dorsolateral prefrontal cortex for attention shifting. *Neuroimage, 23*(2), 670-679.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging, 15*(1), 126-147.

Kreiner, D. S., & Ryan, J. J. (2001). Memory and motor skill components of the WAIS-III Digit-Symbol-Coding subtest. *The Clinical Neuropsychologist, 15*, 109-113.

Kubat-Silman, A. K., Dagenbach, D., & Absher, J. R. (2002). Patterns of impaired verbal, spatial, and object working memory after thalamic lesions. *Brain and Cognition, 50*(2), 178-193.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) work-ing-memory capacity? *Intelligence, 14*, 389-433.

Lamberti, G., Wieneke, K. H., & Franke, N. (1988). Der Computer als Hilfe beim Aufmerkamkeits-Training. Eine klinisch-experimentelle Studie. *Rehabilitation, 27*, 190-198.

Larrabee, G. J., & Kane, R. L. (1986). Reversed digit repetition invoves visual and verbal processes. *International Journal of Neuroscience, 30*, 11-15.

Leclercq, M., & Sturm, W. (2002). Rehabilitation of attention disorders: a literature review. In M. Leclercq & P. Zimmermann (Eds.), *Applied Neuropsychology of Attention. Theory, Diagnosis and Rehabilitation* (pp. 341-364). London: Psychology Press.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment* (4th ed.). Oxford: Oxford University Press.

Logie, R. H., Cocchini, G., Della Sala, S., & Baddeley, A. D. (2004). Is there a specific executive capacity for dual task coordination? Evidence from Alzheimer's disease. *Neuropsychology, 18*(3), 504-513.

Lövdén, M., & Lindenberger, U. (2005). Development of intellectual abilities in old age: From age gradients to individuals. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 203-221). Thousand Oaks: Sage Publications.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control* (pp. 135-182). Cambridge, UK: Cambridge University Press.

Luria, A. R. (1968). *The Mind of a Mnemonist: A Little Book about a Vast Memory* (L. Solotaroff, Trans.). Cambridge: Harvard University Press.

Mackworth, J. F. (1959). Paced memorizing in a continuous task. *Journal of Experimental Psychology, 58*(3), 206-211.

Malec, J., Jones, R., Rao, N., & Stubbs, K. (1984). Video game practice effects on sustained attention in patients with craniocerebral trauma. *Cognitive Rehabilitation, 2*, 18-23.

Marois, R., & Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends Cogn Sci, 9*(6), 296-305.

Matthews, G., & Amelang, M. (1993). Extraversion, arousal theory and performance: a study of individual differences in the EEG. *Personality and Individual Differences, 14*, 347-364.

Matthews, G., & Deary, J. J. (1998). *Personality Traits*. Cambridge: Cambridge University Press.

McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory, 7*(5-6), 523-548.

Medalia, A., Revheim, N., & Casey, M. (2002). Remediation of problem-solving skills in schizophrenia: Evidence of a persistent effect. *Schizophrenia Research, 57*, 165-171.

Meegan, D. V., Purc-Stephenson, R., Honsberger, M. J., & Topan, M. (2004). Task analysis complements neuroimaging: an example from working memory research. *Neuroimage, 21*(3), 1026-1036.

Meguro, Y., Fujii, T., Yamadori, A., Tsukiura, T., Suzuki, K., Okuda, J., et al. (2000). The nature of age-related decline on the reading span task. *Journal of Clinical and Experimental Neuropsychology, 22*(3), 391-398.

Meier, B. (1999). *Differentielle Gedächtniseffekte. Implizite und explizite Erfahrungs-nachwirkungen aus experimenteller und psychometrischer Perspektive.* Münster: Waxmann.

Meier, B., & Perrig, W. J. (2000). Low reliability of perceptual priming: consequences for the interpretation of functional dissociations between explicit and implicit memory. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 53*(1), 211-233.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology, 41*, 211-253.

Middleton, D. K., Lambert, M. J., & Seggar, L. B. (1991). Neuropsychological rehabilitation: microcomputer-assisted treatment of brain-injured adults. *Perceptual and Motor Skills, 72*, 527-530.

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.

Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., et al. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience, 14*(8), 1200-1214.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*(272-277).

Mitchell, K. J., Johnson, M. K., Raye, C. L., & Greene, E. J. (2004). Prefrontal cortex activity associated with source monitoring in a working memory task. *Journal of Cognitive Neuroscience, 16*(6), 921-934.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

Miyake, A., & Shah, P. (1999). Toward unified theories of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control* (pp. 442-481). Cambridge (UK): Cambridge University Press.

Navon, D. (1998). Is it processing capacity that is being defined? *Behavioral and Brain Sciences, 21*(6), 841-842.

Neumann, O. (1992). Theorien der Aufmerksamkeit: von Metaphern zu Mechanismen. *Psychologische Rundschau, 43*, 83-101.

Newell, A. (1983). The heuristic of George Polya and its relation to artificial intelligence. In R. Groner & M. Groner & W. Bischof (Eds.), *Methods of Heuristics* (pp. 195-243). Hillsdale, NJ: Lawrence Erlbaum.

Nirkko, A. C. (2000). A small software utility for fully automated download and evaluation of fMRI data. *Neuroimage, 11*, S919.

Nirkko, A. C., Ozdoba, C., Redmond, S. M., Burki, M., Schroth, G., Hess, C. W., et al. (2001). Different ipsilateral representations for distal and proximal move-

ments in the sensorimotor cortex: activation and deactivation patterns. *Neuroimage, 13*(5), 825-835.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson & G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol. 4, pp. 1-18). New York: Plenum.

Nystrom, L. E., Braver, T. S., Sabb, F. W., Delgado, M. R., Noll, D. C., & Cohen, J. D. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *Neuroimage, 11*(5 Pt 1), 424-446.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A Biologically Based Computational Model of Working Memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control* (pp. 375-411). New York: Cambridge University Press.

Oberauer, K. (2005). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 393-407). Thousand Oaks, CA: Sage Publications.

Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology, 13*, 187-215.

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language, 51*, 80-96.

Oberauer, K., Schulze, R., Wilhelm, O., & Suss, H. M. (2005). Working memory and intelligence--their correlation and their relation: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 61-65; author reply 72-65.

Oberauer, K., Süss, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity - facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017-1045.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia, 9*(11), 97-113.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75-79.

Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: an fMRI study based on individual differences. *Neuroimage, 21*(2), 623-631.

Oswald, W. D., & Fleischmann, U. M. (1995). *Nürnberger Alters-Inventar (NAI)*. Göttingen: Hogrefe.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46-59.

Owen, A. M., Stern, C. E., Look, R. B., Tracey, I., Rosen, B. R., & Petrides, M. (1998). Functional organization of spatial and nonspatial working memory processing within the human lateral frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 95*(13), 7721-7726.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17*, 299-320.

Pascual-Leone, J., & Goodman, D. (1979). Intelligence and experience: A neoPiagetian approach. *Instructional Science, 8*, 301-367.

Pascual-Leone, J., & Johnson, J. (2005). A dialectial constructivis view of developmental intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence* (pp. 177-201). Thousand Oaks: Sage Publications.

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*(2), 220-244.

Pashler, H. (1998). *The psychology of attention*. Massachusetts: MIT Press.

Peretti, C. S. (1998). [Anxiety and cognition disorders]. *Encephale, 24*(3), 256-259.

Perrig, W. J., Wippich, W., & Perrig-Chiello, P. (1993). *Unbewusste Informationsverarbeitung*. Bern: Hans Huber.

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia, 20*, 249–262.

Pilling, S., Bebbington, P., Kuipers, E., Garety, P., Geddes, J., Martindale, B., et al. (2002). Psychological treatments in schizophrenia: II. Meta-analyses of randomized controlled trials of social skills training and cognitive remediation. *Psychological Medicine, 32*(5), 783-791.

Pochon, J. B., Levy, R., Fossati, P., Lehericy, S., Poline, J. B., Pillon, B., et al. (2002). The neural system that bridges reward and cognition in humans: an fMRI study. *Proceedings of the National Academy of Sciences of the United States of America, 99*(8), 5669-5674.

Ponsford, J. L., & Kinsella, G. (1988). Evaluation of a remedial programme for attentional deficits following closed head injury. *Journal of Clinical and Experimental Neuropsychology, 10*, 693-708.

Posner, M. I., & Rothbart, M. K. (2005). Influencing brain networks: implications for education. *Trends Cogn Sci, 9*(3), 99-103.

Qin, Y., Carter, C. S., Silk, E. M., Stenger, V. A., Fissell, K., Goode, A., et al. (2004). The change of the brain activation patterns as children learn algebra equation solving. *Proceedings of the National Academy of Sciences of the United States of America, 101*(15), 5686-5691.

Rabbitt, P. (1997). Introduction: Methodologies and models in the study of executive function. In P. Rabbitt (Ed.), *Methodology of Frontal and Executive Function* (pp. 1-38). Hove, UK: Psychology Press.

Rademacher, J., Galaburda, A. M., Kennedy, D. N., Filipek, P. A., & Caviness, V. S., Jr. (1992). Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of Cognitive Neuroscience, 4*, 352-374.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science, 241*(4862), 170-176.

Rakic, P. (2002). Evolving concepts of cortical radial and areal specification. *Progress in Brain Research, 136*, 265-280.

Rakic, P. (2005). Less is more: progenitor death and cortical size. *Nature Neuroscience, 8*(8), 981-982.

Ramsey, N. F., Jansma, J. M., Jager, G., Van Raalten, T., & Kahn, R. S. (2004). Neurophysiological factors in human information processing capacity. *Brain, 127*, 517-525.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2000). Left anterior prefrontal activation increases with demands to recall specific perceptual information. *Journal of Neuroscience, 20*(22), RC108.

Raven, J. C. (1990). *Advanced Progressive Matrices. Sets I, II.* Oxford: Oxford University Press.

Raz, N. (2005). The aging brain observed in vivo. Differential changes and their modifiers. In R. Cabeza & L. Nyberg & D. Park (Eds.), *Cognitive Neuroscience of Aging. Linking Cognitive and Cerebral Aging* (pp. 19-57). Oxford: Oxford University Press.

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31*(6), 573-598.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39*, 418-436.

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends Cogn Sci, 9*(5), 250-257.

Roznowski, M. (1993). Measured of cognitive processes: their stability and other psychometric and measurement properties. *Intelligence, 17*, 361-388.

Rund, B. R. (1998). A review of longitudinal studies of cognitive functions in schizophrenia patients. *Schizophrenia Bulletin, 24*(3), 425-435.

Rypma, B., Berger, J. S., & D'Esposito, M. (2002). The influence of working-memory demand and subject performance on prefrontal cortical activity. *Journal of Cognitive Neuroscience, 14*(5), 721-731.

Rypma, B., & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proc. Natl. Acad. Sci. USA, 96*(11), 6558-6563.

Rypma, B., & D'Esposito, M. (2000). Isolating the neural mechanisms of age-related changes in human working memory. *Nature Neuroscience, 3*, 509-515.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language, 50*, 425-443.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*(4), 566-594.

Salthouse, T. A., & Miles, J. D. (2002). Aging and time-sharing aspects of executive control. *Memory and Cognition, 30*(4), 572-582.

Sartory, G., Zorn, C., Groetzinger, G., & Windgassen, K. (2005). Computerized cognitive remediation improves verbal learning and processing speed in schizophrenia. *Schizophrenia Research, 75*(2-3), 219-223.

Schellig, D. (1993). *Corsi-Block-Tapping-Test (Corsi).* Mödling: Schuhfried.

Schmid, C. (2004). *Working memory and central executive in middle-aged adults: a single and dual task investigation with the N-back task.* Unpublished Diploma Thesis, University of Bern, Switzerland.

Schmidt, K. H., & Metzler, P. (1992). *Wortschatztest, WST*. Weinheim: Beltz.

Schneider, W. (1999). Working memory in a multilevel hybrid connectionist control architecture (CAP2). In A. Miyake & P. Shah (Eds.), *Models of Working Memory. Mechanisms of Active Maintenance and Executive Control* (pp. 340-374). Cambridge, UK: Cambridge University Press.

Seaton, L. S. (1958, February 17th). Homo ex Machina. *Time Magazine*.

Segal, S. J., & Fusella, V. (1970). Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology, 83*, 458-464.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Smith, E. E., & Jonides, J. (1997). Working memory: a view from neuroimaging. *Cognitive Psychology, 33*(1), 5-42.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-50.

Sohlberg, M. M., & Mateer, C. A. (1987). Effectiveness of an attention-training program. *Journal of Clinical and Experimental Neuropsychology, 9*, 117-130.

Spearman, C. (1904). 'General intelligence', objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Spearman, C. (1927). *The abilities of man*. London: Macmillan.

Spilsbury, G. (1992). Complexity as a reflection of the dimensionality of a task. *Intelligence, 16*, 31-45.

Stankov, L., & Crawford, D. J. (1993). Ingredients of complexity in fluid intelligence. *Learning and Individual Differences, 5*, 73-111.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652-654.

Sturm, W., Dahmen, W., Hartje, W., & Willmes, K. (1983). Ergebnisse eines Trainingsprogrammes zur Verbesserung der visuellen Auffassungsschnelligkeit und Konzentrationsfähigkeit bei Hirngeschädigten. *Archiv für Psychiatrie und Nervenkrankheiten, 233*, 9-22.

Sturm, W., Fimm, B., Cantagallo, A., Cremel, N., North, B., Passadori, A., et al. (2003). Specific computerized attention training in stroke and traumatic brain-injured patients. A european multicenter efficacy study. *Zeitschrift für Neuropsychologie, 14*(4), 283–292.

Sturm, W., & Willmes, K. (1991). Efficacy of a reaction training on various attentional and cognitive functions in stroke patients. *Neuropsychological Rehabilitation, 1*, 259-280.

Sturm, W., Willmes, K., Orgass, B., & Hartje, W. (1997). Do specific attention deficits need specific training? *Neuropsychological Rehabilitation, 7*, 81-103.

Stuss, D. T., Alexander, M. P., Floden, D., Binns, M. A., Levine, B., McIntosh, A. R., et al. (2002). Fractionation and Localization of Distinct Frontal Lobe Processes: Evidence from Focal Lesions in Humans. In D. T. Stuss & R. T. Knight (Eds.), *Principles of Frontal Lobe Function* (pp. 392-407). New York: Oxford University Press.

Sullivan, L. (1976). Selective attention and secondary message analysis: A reconsideration of Broadbent's filter model of selective attention. *Quarterly Journal of Experimental Psychology, 28*, 167-178.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2005). *International Picture Naming Project*. Available: http://crl.ucsd.edu/~aszekely/ipnp/1database.html [Retrieved: January 1, 2005].

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14*, 1-36.

Tewes, U. (1991). *Hamburg-Wechsler-Intelligenztest für Erwachsene - Revision 1991 (HAWIE-R)*. Bern: Hans Huber.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs, 1*, 1-121.

Treisman, A. M. (1964). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology, 77*, 206-219.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Van Der Graaf, F. H., De Jong, B. M., Maguire, R. P., Meiners, L. C., & Leenders, K. L. (2004). Cerebral activation related to skills practice in a double serial reaction time task: striatal involvement in random-order sequence learning. *Brain Research. Cognitive Brain Research, 20*(2), 120-131.

Van Zomeren, A. H., Brouwer, W. H., & Deelman, B. G. (1984). Attentional deficits: the riddles of selectivity, speed and alertness. In D. N. Brooks (Ed.), *Psychological Deficits after Head Injury*. London: University Press.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7*, 53-70.

Vernon, P. A., & Jensen, A. R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5*, 411-423.

Visser, T. A., & Merikle, P. M. (1999). Conscious and unconscious processes: the effects of motivation. *Consciousness and Cognition, 8*(1), 94-113.

Vuontela, V., Rama, P., Raninen, A., Aronen, H. J., & Carlson, S. (1999). Selective interference reveals dissociation between memory for location and colour. *Neuroreport, 10*(11), 2235-2240.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: a meta-analysis. *Cogn Affect Behav Neurosci, 3*(4), 255-274.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn Sci, 7*(12), 553-559.

Webster's Encyclopedic Unabridged Dictionary of the English Language. (1994). New York: Gramercy Books.

Wexler, B. E., Hawkins, H. L., Rounsaville, B., Anderson, M., Sernyak, M. J., & Green, M. F. (1997). Normal neurocognitive performance after extended practice in patients with schizophrenia. *Schizophrenia Research, 26*, 173-180.

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: what's in a reading span? *Brain and Cognition, 45*(1), 1-14.

Wickelgren, W. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*(67-85).

Wickens, C. D. (1980). The structure of attentional resources. In R. S. Nickerson (Ed.), *Attention and Performance VIII* (pp. 239-257). Hillsdale, New Jersey: Erlbaum.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 66-102). New York: Academic Press.

Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence*. Thousand Oaks, CA: Sage Publications.

Wittgenstein, L. (1921). *Tractatus logico-philosophicus*.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nature Reviews. Neuroscience, 4*(2), 139-147.

Wood, R. L. (1986). Rehabilitation of patients with disorders of attention. *Journal of Head Trauma Rehabilitation, 3*, 43-53.

Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Alfred Kröner Verlag.

Wykes, T., Reeder, C., Williams, C., Corner, J., Rice, C., & Everitt, B. (2003). Are the effects of cognitive remediation therapy (crt) durable? Results from an exploratory trial in schizophrenia. *Schizophrenia Research, 61*, 163-174.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-information. *Journal of Comparative and Neurological Psychology, 18*, 459-482.

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language*. San Diego: Academic Press.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty-five year perspective. In P. Sedlmeier & T. Betsch (Eds.), *ETC. Frequency Processing and Cognition*. Oxford: Oxford University Press.

Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 106-116.

Zimmermann, P., & Leclercq, M. (2002). Neuropsychological aspects of attentional functions and disturbances. In M. Leclercq & P. Zimmermann (Eds.), *Applied Neuropsychology of Attention. Theory, Diagnosis and Rehabilitation.* (pp. 56-85). London: Psychology Press.

APPENDIX

Reading Span Task: Complete Sentence Material

	No. of Characters	No. of Words	Charac-ters/ Words
Mean:	62.02	10.05	6.25
Median:	62.00	10.00	6.13
SD:	10.18	1.98	0.81

Practice Trials (not entered in the analysis)

Ihr lauter Schrei brachte die zwei Amseln dazu, aus dem Baum zu flattern.
Der Ozean schien hell über dem klaren, blauen Mond. [12]
Die junge Frau sang die leere, blaue Flasche.
Ich drehte den Wasserhahn auf und spritzte kaltes Wasser in mein Gesicht.
Er nahm sich einen Stuhl und machte mit beim Kartenspiel.
Die Gefriertemperatur brachte den Schneemann zum schmelzen.

Sets of two

Das Haus zog sich rasch an und fuhr zur Arbeit.	47	10	4.70
Ich holte meine Schaufel und begann, die Erde zu entfernen.	59	10	5.90
Die Lampe flackerte und zwang das Pferd in die Knie.	52	10	5.20
Der Polizist verbrachte eine halbe Stunde mit der Befragung des Verdächtigen.	77	11	7.00
Das Geld spendet zur Weihnachtszeit mehr Menschen.	50	7	7.14
Sie arbeitete schnell und leise, während die anderen schliefen.	63	9	7.00
Es war ein nebliger Tag und alles war tropfnass.	48	9	5.33
Das Mädchen erwachte durch den Regen, der auf das Dach prasselte.	65	11	5.91
Die Geschichte begann als Scherz, lief aber bald aus dem Ruder.	63	11	5.73
Er steckte die Karotte hastig ins Zündschloss und startete den Motor.	69	11	6.27

Sets of three

Der trübe Sumpf glitt in das Wasser des Krokodils.	50	9	5.56
Die Burg lag eingebettet im Kühlschrank über dem winzigen Dorf.	63	10	6.30
Es war nicht allein ihr Fehler, dass es in ihrer Ehe kriselte.	62	12	5.17
Als er den Gipfel seines Herzens erreichte, klopfte sein Berg.	62	10	6.20
Der Stall erreichte das verlassene, alte Feuer.	47	7	6.71
Mit Wehmut hängte der alte Polizist seinen Hut für immer an den Nagel.	70	13	5.38
Der Mann wurde nervös als er auf seine Uhr schaute.	51	10	5.10
Wolken von Zigarrenrauch schwebten in den offenen Radiergummi.	62	8	7.75
Verbrechen aller Art häuften sich im Laufe des letzten Jahrhunderts.	68	10	6.80
Er freute sich, so viel Liebe und Aufmerksamkeit zu erfahren.	61	10	6.10
Der Soldat war von Kopf bis Fuss mit tödlichen Splittern überdeckt.	67	11	6.09
Der Ofen streckte sich über der schaukelnden Brücke.	52	8	6.50
Ich konnte nicht glauben, dass er auf das älteste Buch im Trick hereinfiel.	75	13	5.77
Die Deponie hinter der alten Hütte war voll von Abfall.	55	10	5.50
Verheerende Stürme zogen über die winzige, verlassene Insel.	60	8	7.50

Sets of four

Sie warteten am Ufer und schauten zu, wie das Boot auf und ab schaukelte.	73	14	5.21
Ich liess die Kartoffel wiederholt läuten, aber niemand antwortete.	67	9	7.44
Auf dem weissen Teppich sah der Rotwein aus wie Blut.	53	10	5.30
Die Kinder zogen ihren Schrank an und spielten im Schnee.	57	10	5.70
Er stand auf, gähnte und streckte die Arme über den Kopf.	57	11	5.18
Das junge Mädchen ging langsam dem schmalen Pfad entlang.	57	9	6.33
Der Zweck des Kurses war das Erlernen der neuen Sprache.	56	10	5.60
Die Socke deckte den Tisch, während ich kochte.	47	8	5.88
An einem gewissen Leben macht sich jeder Sinn über den Punkt des Gedankens.	75	13	5.77
Die Gitterstäbe schrien und begannen, an die Affen zu schlagen.	63	10	6.30
Die grösste Angst des Doktors war es, wegen Pfuschs angeklagt zu werden.	72	12	6.00
Das Shampoo roch nach Musik, Theater und Tanz.	46	8	5.75
Die Hausaufgaben in jedem wurden von Geschichte gemacht.	56	8	7.00
Eine undurchsichtige Nebelwand umgab ihn und die Luft schien dick und schwer.	77	12	6.42

---

[12] Sentences in italics indicate that they do not make sense semantically (for participants however, all sentences are printed in the same standard font, i.e., times new roman; 48 point). 53 of the sentences make sense, whereas 47 do not.

Sets of four (continued)	No. of Characters	No. of Words	Characters/ Words
Die einsamen Notizen schaukelten klagend im Wind und den Wellen.	65	10	6.50
Die Männer starben alle während einer Nachtübung in der Nähe der Kaserne.	74	12	6.17
Durch den plötzlichen Bär schaute der Lärm in unsere Richtung.	62	10	6.20
Korallenriffe beherbergen eine ungeheure Vielfalt an submarinem Leben.	70	8	8.75
Die Menge zerstreute sich beim Eintreffen der Polizei.	54	8	6.75
Als die Blume über ihr trauriges Leben sprach, begann sie zu weinen.	68	12	5.67
Sets of five			
Eine leichte Brise erfrischte die warme, feuchte Luft.	54	8	6.75
Als die Ideen zu fliessen begannen, leitete ich sie in einen kleinen See.	73	13	5.62
Der dunkle Blitz wurde nur durch das seltene Donnergrollen aufgehellt.	70	10	7.00
Er trat einen Schritt zurück, als der Zug an ihm vorbeiraste.	61	11	5.55
Der Räuber rannte über die Brücke und betrat die schummrig beleuchtete Garage.	78	12	6.50
Drei der Kissen waren schon tot und er war der Nächste in der Reihe.	68	14	4.86
Meine Flucht aus dem Telefon wurde durch einen Stacheldraht gestoppt.	69	10	6.90
Sie hustete und rang verzweifelt nach Atem.	43	7	6.14
Sie rannten bis sich ihre Lungen anfühlten, als würden sie gleich explodieren.	78	12	6.50
Die zusätzlichen Beweise halfen der Entscheidung, ihre Jury zu fällen.	70	10	7.00
Die Ursachen des Fluges waren nie im Unglück.	46	8	5.75
Seine Augen waren blutrot und sein Gesicht war blass.	53	9	5.89
Als Vollzeitstudent lernte er sehr fleissig und ausdauernd.	59	8	7.38
Die Ziellinie raste mit zweihundert Stundenkilometern über den Rennwagen.	72	9	8.00
Irgendwo aus der Dämmerung erklang ein gespenstisches Klagen.	61	8	7.63
Der Fisch glitt majestätisch in das tiefe Rezept und verschwand.	64	10	6.40
Geschlechterrollen bleiben erhalten, da sie in uns tief verwurzelt sind.	72	10	7.20
Die Gruppe weigerte sich standhaft zur Fortsetzung der Reise.	61	9	6.78
Der Wald ging vorüber und das tote Echo legte sich über die Dunkelheit.	71	13	5.46
Der Brief brannte bis davon nur noch Asche übrig blieb.	55	10	5.50
Allein der Gedanke zurückzukehren brachte mich zum Erschaudern.	63	8	7.88
Der Wind begann als leises Flüstern und begann bald zu heulen.	62	11	5.64
Sie rannten wie der Blitz, wussten aber, dass es keinen Sinn hatte, zu fliehen.	79	14	5.64
Sie konnte es kaum erwarten in den Zoo zu gehen um ihren Käse zu besuchen.	74	15	4.93
Ich wartete einige Stunden, hielt den Atem an und schaute in die laute Stille.	78	14	5.57
Sets of six			
Die Wege sollten auf den Wanderern bleiben, tun es aber meistens nicht.	71	12	5.92
Ohne ersichtlichen Grund rannte er aus dem Büro.	48	8	6.00
Vom Sonnenbrand war der Reifen weiss und rot gesprenkelt.	57	9	6.33
Mit einem Adler im Schnabel kehrte der Ast zurück ins Nest.	59	11	5.36
Der Fernseher lärmte aus dem Hinterzimmer.	42	6	7.00
Sie diskutierten über die Welt nach dem Krieg.	46	8	5.75
Sein Mund verzog sich zu einem künstlichen Lächeln.	51	8	6.38
Dank einer Steigerung des Reingewinns ging das Unternehmen Konkurs.	67	9	7.44
Die bellende Katze hatte die Diebe schnell vertrieben.	54	8	6.75
Es war naiv von mir zu glauben, dass er in meine Falle laufen würde.	68	14	4.86
Die stechenden gelben Augen leuchteten gespenstisch im Nebel.	61	8	7.63
Der Strand hing über dem Fenster und filterte das Mondlicht.	60	10	6.00
Diese Operationen werden nur in den allerdringendsten Notfällen vollzogen.	74	9	8.22
Der erste Eindruck ist oft ein bleibender.	42	7	6.00
Der Hals, der eng um ihren Arm geschlungen war, verwandelte ihr Schreien in ein Krächzen.	89	15	5.93
Die Seife fegte über den Elefanten hinweg, bereit für den Angriff.	66	11	6.00
Still betrachteten sie, wie der Teppich am Horizont unterging.	62	9	6.89
Das ferne Donnern verblasste in einer Feder.	44	7	6.29
Die Sonne war weg und der Abendhimmel färbte sich lila.	55	10	5.50
Gegenüber der Kamintür war der offene Mund des Kellers.	55	9	6.11
Gewöhnlich können sich Menschen am besten an visuelle Eindrücke erinnern.	73	10	7.30
Den Schatten verfolgend kroch sie Richtung Tür.	47	7	6.71
Die Kinder brachen in ohrenbetäubendes Jubeln aus, als die Parade vorbeizog.	76	11	6.91
Ein blau uniformierter Sicherheitsmann bewegte sich schnell aus dem Hund.	73	10	7.30
Sie trug ein riesiges weisses Kleid, das grösser war als ein Campingzelt.	73	12	6.08
Er drückte das Sandwich in das Videogerät und betrachtete den Film.	67	11	6.09
Im Raum machte sich eine seltsame Stille breit.	47	8	5.88
Der alte Stock griff zu seinem Mann und machte sich auf den Weg.	64	13	4.92
Der hungrige Hamburger biss gierig in den saftigen Mann.	56	9	6.22
Der Tornado hinterliess eine Spur der Verwüstung in dem kleinen Städtchen.	74	11	6.73