Proportional slowing in old adults is modulated by episodic memory demands

An investigation of age-related slowing using compatible and arbitrary stimulus-response mappings in the Stroop task and related paradigms

Jochen Laubrock
Allgemeine Psychologie I
Universität Potsdam

Dissertation
zur
Erlangung des Doktorgrades
Dr. phil.
im Fach Psychologie

Eingereicht bei der
Humanwissenschaftlichen Fakultät
der Universität Potsdam
im Jahr 2004

To Hannah and Josefine

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation was conducted at Prof. Reinhold Kliegl's cognitive psychology unit at the University of Potsdam and started as part of the project "'Episodische Akkumulatoren' und kognitive Verlangsamung im Alter" (Ma 2154/1-1), principle investigator Dr. Ulrich Mayr. I am grateful to the Deutsche Forschungsgemeinschaft for the financial support of the project, including but not limited to my salary and the subject fees.

First and foremost, I would like to express gratefulness to my mentors Ulrich Mayr and Reinhold Kliegl. I am most grateful to Ulrich Mayr for laying the ground of this work and for his mentorship and excellent supervision especially during the early phases, and continuous support during later stages. His contributions to this dissertation were of the utmost importance. With his sharp intellect and focused attention he helped to shape the theoretical and experimental ideas and to pinpoint potential problems with interpretation of the results. Reinhold Kliegl deserves thanks for creating an atmosphere and organizational setting that leaves little to be desired, for allowing the freedom for creativity to unfold and at the same time engaging control when needed, and for his creativity and enthusiasm about cognitive science in general and experimental design and data analysis in particular that undoubtedly inspired many of the present ideas. It should also be mentioned that he is a great teacher.

I would further like to thank Dr. Ralf Engbert and Prof. Reinhold Kliegl for offering me the opportunity to stay in Potsdam after Ulrich Mayr's leave by employing me in their exciting project on eye movements in reading and attention. This was a great chance for me to gain insight into a new field, combining highly innovative methods of data analysis and computational modelling with sound experimental work. While the project led to some distraction from my dissertation, this was more often welcome than not. On the other hand, I have to apologize for the fact that work on the dissertation sometimes kept me from devoting more resources to the eye movement project.

I am grateful to several visiting scholars, with whom I had the chance to discuss the data on which the present work is based and who provided valuable comments. In particular, I would like to thank Paul Verhaeghen for an in-depth discussion of Brinley analysis, and Akira Miyake as well as Randy Engle for discussions on aspects of working memory and executive control.

A very special thanks goes out to all of my colleagues at the vibrant and highly stimulating Potsdam cognitive psychology unit. Many aspects of the current dissertation were shaped by discussions with Dr. Klaus Oberauer, with whose sharp-wittedness and expertise in the field I continue to be impressed. Ralf Engbert taught me many aspects of data analysis and modelling from a physicists point of view in a very clear style. I would like to thank my present and former colleagues Ralf Th. Krampe, Thomas Liebscher, Martina Junker, Eike Richter (thanks also for LaTeX tips and 'regressive' discussions), Elke Lange, Antje Nuthmann, Martin Rolfs, Robin Hörnig, Andrea Weidenfeld, Thomas Weskott, Katrin Fischer, Kerstin Vockenberg, Doris Philipp, Peter beim Graben, and Johannes Haack for stimulating discussions about experiments in preparation and fresh data, as well as about many topics in cognitive science and other areas (including non-scientific ones).

I especially wish to thank everyone who was involved with data collection and coding. First and foremost, I would like to thank Petra Grüttner. Without her organizational talent, recruiting and testing of the subjects would have taken much longer and likely been less well documented. I also thank André Krügel, Jan Schumann, Anja Felbrich, Sarah Risse, and Christiane Thomas, who assisted during many hours of testing and coding. Last, but not least, I thank the old and young participants in the experiments.

I would further like to express my thanks to the people who provided a reliable backbone of technical and organizational support, including Robert Brenstein, who taught me many programming tricks and from whose deep knowledge about the inside of the Macintosh I profited a lot, Hannelore Gensel, who provided quick technical assistance, and Mary Gaebel and Nicole Stietzel, who knew the answers to all organizational questions.

A work of this size will inevitably contain typos and other errors. For helping to significantly reduce their number I heartily thank Jutta Rodemann, Antje Nuthmann, and especially Else Coenen for their excellent editorial work. Else made more than just editorial contributions. In many discussions, her precise reviews and focused comments helped to clarify muddled conceptual issues and to shape the line of argumentation.

Introduction

When we get older, our intellectual performance tends to decline. The rate of decline seems to be accelerated, so that it is getting more noticeable after the age of 60. What is causing cognitive aging? Is there one general underlying mechanism, or are several processes involved? Is the nature of the mechanism(s) unspecific or rather specific? Does slowing act at a basal level or at a higher level? Research in cognitive aging has been motivated by these questions at least since the early work of Birren, Botwinick, and Welford (Birren & Botwinick, 1955b, 1955a; Birren, 1959; Welford, 1965). One of the dominant positions has been that slowing is general, i.e., task-independent, and that the underlying mechanism is a very basal one that influences all stages of mental processing to a similar extent. The main empirical pattern that inspired the formulation of 'general slowing' theories is the finding that there is a strong positive correlation of age differences and task complexity: the more complex a task, the larger the age effects. Several mechanisms underlying general slowing have been proposed at various levels of abstraction. For example, a prominent psychometric construct involved by general slowing theorists is 'speed of processing', which is assumed to decelerate in old age. At a lower, more neurobiological level, mechanisms such a decrease in neuronal myelinization, or age-related changes in the catecholaminergic neurotransmitter system, especially regarding dopamine, have been suggested. Common to all general slowing theories is the assumption of a basal defect.

Recently, however, the basality of the mechanism has been questioned (Mayr & Kliegl, 1993; Fisk & Fisher, 1994) on the basis of experimental work showing that different task domains lead to different degrees of age-related slowing. The present study tries to add evidence against the dominating view of general slowing caused by a basal defect. It is argued that slowing as measured in the cognitive laboratory might appear 'general' in the eye of the meta-analyst, but that specific processes differ in their susceptibility to aging. According to the present view, age-related performance decrease is caused by specific higher-level cognitive functions, related to episodic memory, working memory, or executive control, which are not usually considered basal, lower-level mechanisms. Slowing nevertheless appears 'general' since these higher-level functions are required by most tasks typically tested in the laboratory. Furthermore, as tasks become more complex, on average the degree of involvement of higher-level functions becomes larger. Realizing this enables one to construct tasks that evoke differential slowing, depending on the degree to which episodic retrieval and executive control of working memory are needed to effectively deal with the task demands.

One problem with this hypothesis is that the construct of 'executive control' is itself rather ill-defined, although some models have been proposed, for example, the SAS model by Norman and Shallice (1986), or the EC-TVA by Logan and Gordon (2001). While there appears to be consensus that prefrontal cortical networks implement executive control functions, the current degree of resolution of neurophysiological theories concerning the functional architecture as well as of cognitive-behavioral theories concerning executive processing functions is rather coarse. However, the study of executive control has become a very active field recently, and some progress has been made.

To the degree that cognitive aging is caused by a decline in executive functioning, as has been suggested for example by West (1996), its study can contribute to theoretical developments in the latter area. In this report, I mainly focus on executive processes that are responsible for keeping a set of arbitrary task rules active in and coordinating retrieval from episodic long-term memory. The executive aspect of this requires switching the focus of attention as well as protecting memory from stimulus-elicited interference. Both the executive functions of focus switching and interference protection can be thought of as controlled attention processes that are at the heart of working memory function (e.g., Oberauer & Kliegl, 2001). Both processes might also be required to perform well in a task-switching paradigm, the study of which has dominated research in the area of executive control in the last few years. However, the current investigation limits itself to simpler reaction time paradigms, which more closely resemble the tasks that originally generated the database on which general slowing theories rest.

Organization of the thesis

The general argument I propose in this thesis is that although slowing may appear to be caused by a very basal mechanism, it is indeed likely caused a specific deficit in a rather high-level system that is needed for the representation and maintenance of arbitrary task rules. I will first discuss the "complexity effect" (Cerella, Poon, & Williams, 1980; Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Salthouse, 1985), which serves as a data basis for general slowing theories. I will then proceed to discuss some problems with the general slowing approach. These fall into two broad categories. First, complexity and the related notions of cognitive resources and mental speed are not well specified. Second, there is empirical evidence that a general slowing model overestimates age effects in some task domains, and underestimates age effects in other task domains.

Thus, an alternative account of age-related slowing is needed that can account for the complexity effect without the assumption of a basal deficit. The 'Episodic Accumulator Model' is introduced as such an account, based on the observation that task complexity is usually related if not identical to the (episodic and working) memory and controlled attention demands of a task. I propose that age-related slowing originates from defects in specific, high-level processes, which are required to establish and maintain a 'mental set' that specifies arbitrary rules for a given task, i.e., associations between task elements that are not over-learned. Why is the account labeled 'Episodic Accumulator Model'? The processes used to establish and maintain a mental set are similar to the processes commonly discussed under the label episodic memory, because arbitrary task rules vary from task to task and from context to context. An example is the set of stimulus-response (S-R) mappings in a given task, which is in principle the same as an admittedly short paired associate list. Paired associate lists are the stimuli typically used to study episodic memory. Arbitrary task rules often relate internal representations to a required output, hence they come into play relatively late in the cascade of information processes, at the stage of response translation or response selection. Response selection is typically conceived of as taking place in a set of buffers: once the difference in activation between element buffers exceeds a threshold, a response program is initiated (e.g., Hanes & Schall, 1996). Thus taken together, an episodic accumulator is a flexibly instructible buffer at the stage of response translation or selection. The critical aspect that is proposed to be a major determinant of age-related slowing is a loss of reliability of these episodic accumulators. Next, a series of experiments designed to test the episodic buffer account is presented. Common to all experiments are the assumptions that (a) episodic buffers are particularly vulnerable to age effects, and (b) it is possible to create conditions under which episodic buffers can be bypassed. The first experiment uses a perceptual difficulty manipulation in combination with a stimulus-response compatibility manipulation to establish that the locus of age effects is likely to be in a relatively late stage of processing. Because a host of findings from the additive factors research program indicate that manipulations which selectively affect sensorimotor and cognitive stages are largely additive in their effects, Experiment 1 is similar to a "sequential complexity" manipulation (Mayr & Kliegl, 1993; Mayr, Kliegl, & Krampe, 1996; Verhaeghen, Kliegl, & Mayr, 1997), in which later processing steps are relatively independent of the results of earlier processing steps. In terms of the episodic buffer model, this means that manipulations affecting perceptual classification mainly introduce a time delay, after which similar information about stimulus identity is fed into episodic buffers. A second set of experiments using the Stroop task investigates situations in which earlier and later processing are less independent, and hence more "coordinatively complex" processing is required even in superficially simple tasks. These tasks were designed to shift the difficulty manipulations closer to the hypothesized locus of the episodic buffer stage, so that a more cascaded mode of processing is instantiated. According to the Episodic Accumulator Model, under these conditions early cognitive difficulty effects are expected to be age-differentially amplified, compared to conditions in which episodic buffers can be bypassed. The diagnostic data pattern supporting the Episodic Accumulator Model in cascaded mode is therefore an over-additive three-way interaction of age, early cognitive difficulty, and episodic difficulty. Finally, a Brinley plot meta-analysis of the experiments is presented to show that the differential effect of mental sets on age-related slowing observed within single experiments can be confirmed in an analysis across experiments. As a methodological contribution, a multiple-intercept regression model tailored to the processes under investigation is developed. The model is rather successfully applied to the experimental data and shown to be superior to classical Brinley analyses.

Cognitive Aging: Data and Theories

In this section, theories aimed at explaining the complexity effect, which is the central empirical finding in the cognitive aging literature, are presented. These theories have in common that they assume a deficit at a low and unspecific level. The presentation is followed by a discussion of findings that cast doubt on the basality of the slowing mechanism.

General Slowing and the complexity effect

The dominant empirical pattern in cognitive aging is the finding that relatively independently of the type of task, old adults perform worse than young adults. This has led to the formulation of 'general slowing' models, which regard cognitive aging as an unspecific process. They are supported by psychometric results showing that the degree of age effects is highly correlated across tasks (Lindenberger, Mayr, & Kliegl, 1993; Salthouse, 1996), as well as by the so-called "complexity effect". The complexity effect was first observed by Brinley (1965), describing the result that absolute age effects increase with the degree of task complexity, whereas proportional age effects remain relatively constant. Old and young reaction times seem to be related by a constant proportion. The scatter plot of old versus young adults' condition means from a variety of experimental conditions can very well be approximated by a simple linear regression line ('Brinley function') with a slope greater than one. Even more interesting is the meta-analytical result that the slope of the Brinley function is more or less constant, relatively independent of task domain (Cerella, 1985, 1990; Hale & Myerson, 1996). This is a rather stable result, which has led to the formulation of general slowing theories that try to capture the essence of aging in a small number of parameters. One of the earliest and simplest proposals was made by Brinley (1965), who suggested that an aging theory could be regarded as a function that captures the relation between RT_{young} and RT_{old}:

$$RT_{old} = f(task, RT_{young}) \quad . \qquad (1)$$

Given a set of information processing tasks for which reaction times at age 20 and age 70 are known, the theoretical goal is to find a function that describes the performance of the elderly group given the performance of the young group and the requirements of the task. Salthouse (1978) and Cerella et al. (1980) published cross-sectional data sets from a variety of tasks that were well described by a linear function:

$$RT_{old} = \lambda RT_{young} \quad . \tag{2}$$

If we compare the empirically derived equation (2) with the theoretical equation (1), we note that task type does not enter the equation, although the data used in (2) derive from a heterogenous set of tasks. This has been interpreted to indicate that slowing is general in the sense that it does not differentially affect specific processes. Later theoretical developments tried to investigate mechanisms that could lead to such a generalized slowing function. A prominent example is the family of models described by Cerella (1985, 1990), which was explicitly inspired by two key ideas (Cerella, 1990, pp. 201-202)

> One was the realization [...] that age deficits could be interpreted as being distributed throughout the information-processing system rather than being localized in particular stages. References to task content could thereby be eliminated; deficits were tied to the amount, not the type, of the information processing. [...] The second key idea [...] is the attempt to interpret cognition as a computation on a neural network, rather than as a succession of information-processing stages. This idea combines with the preceding idea in a natural way: The new aging theories view age deficits as defects of some sort distributed throughout a neural network of some sort.

The models were developed starting with a simple one that could nevertheless predict some of the classic effects of aging on information processing latencies. In this simple model, the brain is considered to be a (feed-forward) neural network composed of links and nodes. A cognitive process is the propagation of a signal from the input end of the network to the output end. Each step takes a fixed amount of time. Reaction time (RT) is given as the number N of links to be traversed times the time per link, μ. The aging process breaks links in the networks (at a constant rate k), thus requiring detours from a straight path, so that more links have to be traversed. At a fixed age, a fixed proportion p of links will be intact, and the inverse proportion 1-p will be broken. Diversions can lead to further diversions if a broken link is encountered during the diversion. Thus

$$\begin{aligned} RT_{old} &= \mu N + (1-p)\mu N + (1-p)^2 \mu N + \cdots \\ &= \frac{1}{p}\mu N = \frac{1}{p} RT_{young} \quad . \end{aligned} \tag{3}$$

This simple model predicts that latencies of a degraded network will be a constant multiple of the latency of the intact network, regardless of the cognitive processes involved. Thus linear Brinley plots are predicted independent of task domain. In agreement with the data, the model also predicts a result observed in longitudinal studies[1], namely an exponential decline of functionality with age:

$$RT_{old} = exp(ka)\, RT_{young} \quad , \tag{4}$$

because the proportion of intact links is a negatively accelerated function of network age (a).[2]

Later it was realized that even the data used in the early meta-analyses (Cerella et al., 1980) seemed to be more compatible with a multilayer aging model that assumed different slowing factors for peripheral and central stages. In a re-analysis of the 1980 data, Cerella (1985) fitted individual regression lines for each of the 14 experiments (each experiment provided data from several conditions). The resulting lines did not intersect at $[RT_{young}, RT_{old}] = [0,0]$, but at $[RT_{young}, RT_{old}] = [464, 568]$, so that the complete fan of lines was described by

$$(RT_{old} - 568) = \lambda_{exp}(RT_{young} - 464) \quad , \tag{5}$$

where λ_{exp} indicated that the slope was different for each experiment (median slope was 1.82). Cerella (1985) noted that the non-zero intersection point could be accounted for by assuming two separate slowing factors in an additive-stage like model, so that age effects in peripheral, sensori-motor stages are less severe than in a central, cognitive stage. The model is given by

$$\begin{aligned} RT_{young} &= S + C \\ RT_{old} &= \lambda_s S + \lambda_c C \quad . \end{aligned} \tag{6}$$

To see how the empirical fan of lines described by (5) can be predicted by this model, we express C by $RT_{young} - S$ and rearrange terms so that

$$(RT_{old} - \lambda_s S) = \lambda_c\, (RT_{young} - S) \quad . \tag{7}$$

The sensorimotor slowing factor is given by $\lambda_s = 568/464 = 1.22$, and the cognitive slowing factor for a given experiment by the regression slope, λ_c. In this model, cognitive slowing is 'general' as long as λ_c is not allowed to vary between experiments. Data support a general slowing model if the λ_{exp} are not significantly different from each other.

Yet another class of theories of cognitive aging were developed to explain data that are sometimes observed when very heterogeneous conditions are compared in a Brinley plot (e.g., Hale, Myerson, & Wagstaff, 1987). If these data are plotted together with low- and intermediate-complexity conditions, conditions with a

---

[1] In contrast, in the current publication, as well as in most published studies, the focus will be on age effects at cross-section, i.e., the performances of a group of young adults in their twenties and a group of old adults at the age of about 65-75 will be compared.

[2] This simply results from the assumptions of a constant 'decay rate' k, so that $\frac{d}{dt}p(t) = -kp(t)$, from which $p(a) = exp(-ka)$ is obtained by integration, assuming an intact network at the young reference age. If $p(a)$ is substituted for p in equation (3), then (4) is obtained.

very high degree of complexity can produce points in Brinley space that produce positive residuals from a linear regression—i.e. the age difference is larger than expected by even a proportional slowing account like (6). Empirically, Brinley plots of this type were found to be fitted best by a power function, $RT_{old} = \beta\,(RT_{young})^{\lambda}$. To explain this fact, Myerson et al. (1990) have developed the "information loss model", a processing model that assumes that at each micro-step in a series of computations, a certain proportion p of information is lost, that this proportion is larger for old adults than for young adults, and that step duration is inversely proportional to the amount of information available. This model predicts a positively accelerated power function in Brinley space, with an exponent of $\lambda = p_{old}/p_{young}$. The information-loss model is a general slowing model, because slowing is captured in a single parameter.[3]

Theories of general slowing differ with respect to the specificity of the predictions they make. While the models discussed by Cerella (1990) make rather specific assumptions and predictions, thereby allowing for empirical falsification, the processing-speed theory proposed by Salthouse (1996) is formally less well-specified. In particular, despite its name, the concept of "processing-speed" seems to be an umbrella construct that encompasses a relatively large number of cognitive processes. What Salthouse (1996) does claim is that "a small number of common factors contribute to the age-related differences in many speed measures." Processing-speed is postulated to represent how quickly many different types of processing operations can be carried out. It is usually operationalized by psychometric tests of "perceptual speed", such as the Digit Symbol Substitution scale of the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981). Thus any cognitive process or stage that is a component of the speed measures could be responsible for the effect, one candidate mechanism being working memory. However, the label "processing-speed" suggests a more basal mechanism, and indeed, Salthouse and colleagues (Salthouse, 1994, 1991) claim that processing-speed can explain age-related variance better than working memory. The question remains what constitutes processing-speed. One possible answer could be obtained by performing a task-analysis of the tasks typically used to measure processing-speed. One important observation by Salthouse is that "the proportional attenuation of the age-related variance was greater with speed measures from tasks involving perceptual or cognitive operations, such as substitution, transformation, or comparison, than with tasks merely requiring copying or line drawing responses" (Salthouse, 1996, p.420). Thus processing-speed measures central rather then sensorimotor processing. Salthouse also sees similarities between the construct of processing speed and "aspects of attention such as inhibition" or "processing resources". In his opinion, these are less well operationalized than the processing-speed construct.

While he acknowledges that there is little knowledge about the function of mechanisms that relate processing-speed to levels of cognitive performance, Salthouse (1996, p. 425, quoting Salthouse, 1992, p. 116) speculates that the neurophysiological basis could be

> a slower speed of transmission along single (e.g. loss of myelination) or multiple (e.g. loss of functional cells dictating circuitous linkages) pathways, or [...] delayed propagation at the connections between neural units (e.g. impairment in functioning of neurotransmitters, reduced synchronization of activation patterns).

Thus some commitment is made to the possibility of a single basal mechanism.

Recently, Li and colleagues (Li, Lindenberger, & Frensch, 2000; Li, Lindenberger, & Sikström, 2001) have suggested that a general slowing mechanism at the neurobiological level might have the potential to explain (apparently) process-specific slowing at the information-processing level as well as findings such as age differences in fluid intelligence at the psychometric level. They noticed growing evidence for a decline of the functioning of monoamine neurotransmitter systems with age, in particular the dopaminergic system. Li et al. (2001, p.483) propose that age-related declines in dopaminergic activity cause a decrease in the signal-to-noise ratio of neural information processing. They model this effect by using a smaller gain of the sigmoid activation function in the units of a neural network.[4] Comparing simulated 'young' and 'old' networks, they found that the internal stimulus representations of the old network were less distinctive, and that reduced representational distinctiveness led to extensive activation overlap between modules. Li et al. think that "this effect implies that, as people age, mental representations of various events and the contexts within which the events occurred, such as the conversation held with different individuals within a day in different social settings, become less distinct and more confusable with each other." Another recent model inspired by the observation of age-related changes in dopaminergic system activity was proposed by J. Cohen's group (Braver et al., 2001). In this model, the dopaminergic decline is more specifically related to prefrontal cortex functioning, hence specific deficits in context processing are predicted. However, disturbances in the processing of context "impair cognitive control function across multiple domains, including attention, inhibition, and working memory" (Braver et al., 2001, p.746). A specific,

---

[3] The authors have since noted that the model might be too general, because they found empirical evidence for different slowing factors in lexical and nonlexical task domains (Lima, Hale, & Myerson, 1991; Hale, Myerson, Faust, & Fristoe, 1995; Hale & Myerson, 1996; Jenkins, Myerson, Joerding, & Hale, 2000). Thus now different p's are allowed for broad task domains, making the model a 'broad-domain specific' general slowing model.

[4] Typically, an S-shaped function such as the logistic function is used as the activation function. To change the gain of the activation function is (in my intuition) similar to assuming a lower drift rate in diffusion models of information processing (Ratcliff, 1978).

albeit basal, deficit could thus be a cause for the general slowing pattern, because functions needed in many tasks are affected by the deficit. The finding of a single slope in the Brinley plot is thus entirely compatible with the slowing of specific processes, as has also been shown in simulations studying selective age effects on specific task components (see the discussion in the Journals of Gerontology, 49B: Perfect, 1994; Cerella, 1994; Myerson, Wagstaff, & Hale, 1994; Fisk & Fisher, 1994).

To further complicate the matter, while the finding of a single slope does not rule out a specific deficit, the converse could also apply: it might well be possible that a basal neurobiological defect is expressed as a task- or process-specific deficit at the cognitive level. The empirical finding of domain-specific slopes in Brinley plots does not entirely rule out the existence of a single, general, basal defect. However, such a finding does shed light on the type of processing in a task, because domain-specific slopes are not compatible with a serial, consecutive-stage view of processing if a basal defect is assumed (Dunn & Kirsner, 1988). Instead, recurrent processing (with reentrant mechanisms) is required, such as in the Braver et al. (2001) model of DLPFC-ACC interactions in interference resolution, or in current models of working memory.

Problems with General Slowing

In the following sections I will summarize theoretical and empirical problems with general slowing. The main theoretical criticism relates to the fact that general slowing theories originate from a cognitive resource conception, and that the concept of a resource might be too vague to be of much use in theorizing. The empirical criticism stems from observations that a number of observations have been reported that seem difficult to reconcile with a general slowing model like Cerella's multilayer slowing model (equations 5-7). More specifically, there are domains in which the empirical data suggest a lack of age-related slowing, and other domains in which the model clearly underestimates age effects.

Resources, mental speed, and task-complexity are ill-defined

The complexity effect is often interpreted within a cognitive resources framework, where limited-capacity resources, which decline with age, are assumed to determine performance (Navon & Gopher, 1979; Gopher & Navon, 1980; Salthouse, 1988; Kahneman, 1973). The more complex a task is, the more resources are needed. In general, resource theories assume that an age-related decline in cognitive resources is one of the prime causal factors of cognitive aging. One of the prominent models in cognitive aging (Salthouse, 1996) regards the speed of elementary cognitive operations as the major resource related to cognitive aging. Although Salthouse considers speed to be a better-defined construct than attention or working memory, in my opinion, the tests used to operationalize processing-speed tap a variety of cognitive functions, and the specification and consideration of theories of working memory and controlled attention might provide a clearer view on the microstructure of aging. The concept of 'cognitive speed' (or mental speed) is not very well-defined. The concept is popular in differential psychology, where the result of interest is the consistent correlation of measures of processing speed obtained in relatively simple reaction time tasks with intelligence[5], which is usually thought to measure higher-level cognitive abilities (Jensen, 1998; Lehrl & Fischer, 1988). To experimental psychologists, the tasks used to measure "elementary cognitive operations" might not appear quite as elementary as to differential psychologists—for example the tasks might involve choice reaction time, which in itself is the result of a number of component processes (see for example Usher & McClelland, 2001). Another example for a rather complex measure of elementary cognitive operations from the differential psychology tradition is the Digit Symbol Substitution (DSS) test, which is routinely applied in cognitive aging labs to measure "perceptual speed". To the experimental psychologist, the label 'perceptual speed' suggests a truly low-level measure, excluding higher-level cognitive operations. However, good performance in the DSS is certainly aided by working memory capacity. The test requires writing down the associated digit to each element of a series of arbitrary symbols under speed pressure, and the 'code', i.e. the mapping of symbols to digits, is presented at the top of the answer sheet. Performance in this task is aided by the encoding and short-term storage of the digit-symbol associations, because if a given association cannot be retrieved from memory, saccades have to be made from the current symbol position to the 'code' line and back. Even in the latter case subjects will benefit from a well-functioning short-term memory, because it will reduce the average search time for the matching symbol in the code line.

Despite of the often rather complex tasks used to obtain reaction time measures, some intelligence theorists claim that the consistently observed correlation between reaction or inspection time and intelligence is caused by nerve conduction velocity (e.g., Miller, 1994). This is backed by relatively little direct, experimental evidence. Note, however, that sometimes even correlations of visual evoked potentials—presumably truly indicating perceptual speed—and intelligence were reported (Reed & Jensen, 1992).

The conception of cognitive speed—the speed of elementary mental operations—as a resource appears to differ rather radically from the 'cognitive resource' concept discussed in the human perception and performance literature, where cognitive resources have typically been used as a metaphor for attention (Moray, 1967; Kahneman, 1973; Norman & Bobrow, 1975; Wickens, 1984; Wickens & Liu, 1988). How do these apparently heterogenous concepts relate?

---

[5] Here intelligence refers to a construct akin to fluid general ability, Gf, in Cattell's theory, or the general Factor g in Spearman's theory. Jensen (2000) considers these to be the same, Gf=g.

Common to all 'attention as resource' theories is the concept of a limited pool of resources that can be divided between tasks or processes. Broadly, single-resource theories (Moray, 1967; Kahneman, 1973) assume an undifferentiated pool of resources that is available to all tasks. As a task is made more difficult, an increase in the supply of resources is demanded by physiological mechanisms. If the increase is insufficient, performance falls off. However, single resource theories cannot account for all of the results obtained in dual-task interference studies. For example, Shah and Miyake (1996) report that their experiment 'provides preliminary evidence for separate pools of cognitive resources for the two working memories (one for spatial thinking and the other for language processing)'. Wickens' (1984; 1988) multiple resource model acknowledges that different codes have different potential to interfere. Resources are defined by three dimensions: stage (early vs. late), modality (e.g. auditory vs. visual), and processing code (e.g. spatial vs. verbal encoding). When two tasks demand separate resources, efficient time-sharing and little to no interference is expected. Empirical results appear to be more compatible with Wickens' cubic multiple resource model, which of course is not as parsimonious as the single-resource models.

Generally, resource theories have been criticized for their lack of content, and in fact there is no clear-cut, agreed-upon definition for cognitive resources (e.g., Navon, 1984; Pashler, 1998; Oberauer & Kliegl, 2001). One problem with multiple-resource approaches has been to characterize the nature of the resources a particular task is expected to use. Thus, resource theories can be portrayed as rather abstract frameworks that need to be filled with theoretical content to generate useful predictions. Theories such as Baddeley's (1986, 1992a, 2000) working memory model or Pashler's (1998) theoretical framework of attention can be regarded as attempts to clarify the resource concept. For example, in Baddeley's model, the central executive or attention controller component instantiates a mechanism for the scheduling of resources, and slave systems are used for the code-specific storage of phonological and spatial traces.

Not only is there a lack of theoretical grounding for the resource concept, but task complexity, the main theoretical variable determining requirements for cognitive resources, is equally hard to define. For example, various task variations have been suggested as manipulations of task complexity, such as the number of repetitive processing steps, the number of differential mental operations in general, the requirement for parallel processes, or the degree to which complex algorithms have to be assembled. To relate task complexity to resource demands, it is desirable to have a model of the task specifying the involved component processes. While this may be available for some tasks and some experimental manipulations, it is more often not, or at least not made explicit. Integration of data from several tasks in a Brinley plot makes this a more severe problem—in fact, meta-analysis might just pick out the processes that are common to all the included experiments.

Task complexity measures are of central interest to the applied cognitive discipline of Human Factors, where they are usually obtained only after a laborious detailed task analysis. Similarly, within cognitive psychology proper, some attempts have been made to construct task complexity metrics. However, these metrics are usually only applicable to well-defined task domains, e.g. the General Problem Solver (GPS, Newell & Simon, 1961/1963), or the "theory of relational complexity" developed by Halford and colleagues (Halford, Wilson, & Phillips, 1998) as a metric of complexity in reasoning tasks. Halford et al.'s metric proposes that, in addition to the number of unique entities that can be processed in parallel, the structure of the relations between these entities is essential in determining processing capacity limitations. This definition shows that the quantification of task complexity is rather difficult. The fact that many studies in cognitive aging simply use an operational definition lacking specificity—task complexity is equal to performance of young subjects—could be one of the reasons why Brinley analyses have been criticized to obscure specific age-related effects (e.g., Fisk & Fisher, 1994; Perfect, 1994; Myerson et al., 1994).

Coming back to our initial question, are the concepts of cognitive speed as a resource and attention as a resource related? Can controlled attention and speed of processing be equated? If so, then speed of processing is certainly not a basal, low-level mechanism that can be reduced to, for example, nerve conduction speed. If not, then the labels for the capacity-limited resource concepts employed by the two approaches suggest differences where there are none. We draw the preliminary conclusion that the resource and mental speed metaphors should be filled with structural and processing content. Either mental speed is a low-level construct related to basal mechanisms such as nerve conduction speed, or it is a high-level construct related to or even congruent with working memory and controlled attention. In the former case, the finding of different slowing functions for different task domains appears to be problematic for general slowing theories. In the latter case it might make more sense to base theories of aging on models of cognitive processes such as working memory or executive control. Of course most of these processes are theoretical constructs as well, but for these, there exist (in my view) better-specified theories than for a high-level speed construct.

Independent of its theoretical status, there are empirical problems with the general slowing model. General slowing, at least in a serial information processing framework, only allows for a single slowing factor, which is often estimated by the slope of a typical Brinley function to be about 1.5 to 1.8. Thus, results showing different degrees of age-related slowing for different task-domains constitute evidence against general slowing.[6] There is empirical evidence showing

---

[6] This evidence might not be completely decisive, because processing models can be constructed that predict different

that there are task domains where slowing is apparently absent, as well as task domains where the slope is larger than predicted by general slowing. To the degree that the tasks cover a sufficient range of young adults' reaction times, these results are not only problematic for linear regression-type general slowing models, but also for models that predict a power-law relation. The following sections will summarize empirical deviations from the general slowing pattern.

Overestimation of age effects

In a variety of tasks and measures, age-related slowing is virtually absent. In particular, this seems to be the case in tasks that rely on semantic memory (or crystallized intelligence). A recent meta-analysis of aging and vocabulary scores shows that old adults perform better than expected by general slowing in production tests and especially in multiple-choice tests (Verhaeghen, Steitz, Sliwinski, & Cerella, 2003). Furthermore, semantic priming effects are largely age-invariant (e.g. Laver & Burke, 1993; Madden, Pierce, & Allen, 1993; although the reliability of the measures has been questioned, Salthouse, 1996). Similarly, age effects are small in more complex semantic tasks like semantic production, or fluency tasks (Mayr & Kliegl, 2000; Baeckman & Nilsson, 1996; Obler & Albert, 1985; Schaie & Parham, 1977; for a review see Light, LaVoie, Valencia-Laver, Owens, et al., 1992). The small age effects remaining in these tasks can probably be attributed to nonsemantic, e.g. executive or motor, processes contributing to the production time (Mayr & Kliegl, 2000). Taken together, this seems to indicate that semantic long-term memory is less susceptible to age-related degradation than other cognitive processes.

In a recent meta-analysis of negative priming tasks (Verhaeghen, Vandenbroucke, & Dierckx, 1998), the slope of the Brinley function was 1.04, i.e., close to one. This can also be explained by almost non-existing age effects in the semantic domain, because the tasks typically used to measure negative priming require the naming of objects. Hence semantic knowledge is the major determinant of reaction time, in comparison to which the actual negative priming effect is very small.

Mental arithmetic constitutes another domain with small age effects, as long as the executive demands are low. The slowing factor obtained in simple addition and subtraction tasks is only about 1.2 (e.g., Charness & Campbell, 1988; Geary, Frensch, & Wiley, 1993). If peripheral aspects are eliminated by using the method of time-accuracy functions, then slowing is completely absent (Verhaeghen et al., 1997).

When simple reaction time measures are used, 'cognitive' choice components are eliminated from the task. Under these conditions, one can observe age equivalence even in reaction time tasks (Nebes, 1978; Salthouse & Somberg, 1982; Thomas, Waugh, & Fozard, 1978), at least when responding is vocal. Age equivalence is also observed in tasks involving more complex motor programs, as long as the task is highly overlearned. For example, word-naming latency does not increase with age, whereas color-naming latency does (Cohn, Dustman, & Bradford, 1984). More generally, it has been proposed that existing automatic processes are relatively unaffected by the aging process, whereas controlled processes decline with advancing age (Fisk & Rogers, 1991).

Age effects in the verbal domain are also relatively small (e.g., Hale & Myerson, 1996; Lawrence, Myerson, & Hale, 1998; Lima et al., 1991; Jenkins et al., 2000). For example, Hale and Myerson (1996) found that slowing factors in four speeded verbal tasks (lexical decision, double lexical decision, category membership, and synonym-antonym judgment) were rather small compared to a set of visuo-spatial tasks. Jenkins et al. (2000) replicated this result, with an estimated Brinley slope of 1.22 for the verbal domain (vs. 2.56 in the spatial domain), and extended it to show that verbal working memory tasks as well as the learning of novel verbal information produced smaller age effects than the corresponding tasks using spatial material.

While these verbal tasks can still be considered to mainly tap semantic knowledge, there are also domains where the latter plays no role, but age effects are nevertheless small. One example are tasks relying mainly on bottom-up, exogenously driven visuo-spatial attention. Exogenously driven attention is automatically attracted by the stimulus, without an intentional act, and does not draw cognitive 'resources'. For example, Greenwood and colleagues (Greenwood & Parasuraman, 1994; Greenwood, Parasuraman, & Haxby, 1993) found that effects of peripheral cues that attract attention exogenously do not differ as a function of age (see also Folk, Remington, & Johnston, 1992; Gottlob & Madden, 1998).

Similarly, age differences are small in visual search. In a prototypical visual search task, participants have to look for a target amongst a number of distractors. The target is defined either by a simple feature (e.g., green object among red objects; feature search) or by a conjunction of features (e.g., green circle among red and green circles and squares; conjunction search). A dependent variable commonly used in visual search tasks is the slope of the search function, which is a linear function relating search time to the number of distractors. While the search function is flat for feature search, which is interpreted as indicating 'preattentive' parallel search, search time is positively correlated with the number of distractors for conjunction search, leading to a positive search slope. The increase in search time is supposed to indicate the involvement of top-down, controlled attention, which is thought to be either deployed

---

slopes in spite of a single slowing mechanism (Salthouse, 1985). A scenario can be constructed where some closed-loop, recurrent, 'reverbatory circuits' type of information processing is needed in task domain A, but not in task domain B. Thus even if the low-level processing speed account cannot be defeated by such a 'differential slowing' finding, speed theorists would be required to specify the type of information processing needed in a task domain. In this case, at least the data are valuable because they help to specify the type of processing required by a task.

serially to different areas of the display, or in parallel, but with a limited capacity. In a study of age differences in visual search, the feature search slope was flat for both age groups, but, more interestingly, age effects were also rather small in the search slope for conjunction search, although it involves controlled attention (Humphrey & Kramer, 1997). Averaged across (conjunction search) conditions, the slope was 33 ms per item for young adults and 38 ms for old adults, corresponding to a slowing factor of 1.15, which is certainly lower than the typical slope of 1.5-2.0.

Finally, evidence from the on-line monitoring of cognitive processing using event-related potentials (ERP) indicates that the duration of early cognitive processes (possibly corresponding to pre-attentive processing) seems to be unaffected by the aging process. Of particular interest is the result of a meta-analysis comparing age effects on P300 latency (Bashore, Osman, & Heffley, 1989), which is an ERP component that is commonly thought to reflect completion of stimulus evaluation (e.g., McCarthy & Donchin, 1981). The meta-analysis found that P300 latency does not change with age (see however Lubbe & Verleger, 2002); in fact, the Brinley slope for P300 latency was 0.95. In the same data, the slope for reaction times was 1.3. Thus, the phenomenon of 'general slowing' seems to begin only after stimulus evaluation. In later studies it was found that the effect of age on P300 latency was enlarged if the stimuli were degraded, but not if the S-R mapping was incompatible (Smulders, Kenemans, Schmidt, & Kok, 1999), indicating (a) some age-related slowing in the stimulus evaluation stage under perceptually difficult conditions, and (b) an independence of P300 latency from response translation and selection processes.

In summary, age effects appear to be relatively small in tasks where processing either relies on overlearned associations or is otherwise automatic, rather than requiring intervention of controlled attention. This appears to be difficult to explain by variants of the general slowing account that assume a very basal deficit as the cause of age-related slowing.

Underestimation of age effects

The slope of the Brinley function in other task domains is much larger than 1.5. In particular, this seems to be the case when the tasks are 'coordinatively complex'. Coordinative complexity means that earlier cognitive processes provide internal working memory representations that are operated upon by later processes. Furthermore, coordinatively complex tasks are characterized by the fact that flow of information is not strictly serial, but simultaneous storage and processing is required, because intermediate results have to be stored for later retrieval while processing that operates on the same mental code takes place. An example using the task of mental arithmetic will clarify the distinction between sequential complexity and coordinative complexity introduced by Kliegl and colleagues (Mayr & Kliegl, 1993; Mayr et al., 1996; Verhaeghen et al., 1997). Sequential complexity in the mental arithmetic task can be manipulated by directly changing the number of elementary addition and subtraction operations in an arithmetic task, e.g., $6-2+1+4-5-2+1$. In the coordinatively complex version of the task, brackets are introduced, giving e.g., $[6-(2+1)+4]-[5-(2+1)]$, so that the temporary swapping of elements in and out of an intermediate store and the updating of stored elements is required. The task can only be solved if the representations that are swapped into the intermediate storage system remain intact. The Brinley slope in coordinatively complex tasks can be as high as 3.5 to 4.0 (Mayr & Kliegl, 1993; see also Charness & Campbell, 1988; Kliegl, Mayr, & Krampe, 1994).

Very large age effects are also observed in mental rotation using complex stimuli (e.g., Hertzog & Rypma, 1991; Just & Carpenter, 1985; Lohman, 1988). For example, the data points in the study by Hale et al. (1987) that deviated from a linear Brinley function with a pretty much standard slope and that motivated Myerson et al. (1990) to develop the information-loss model all originated from a mental rotation experiment. Mental rotation can be regarded as a coordinatively complex task, because it involves continuous transformation of an object, i.e. earlier representations need to be updated.

Other domains and tasks where large age-effects are commonly observed include divided attention, as indicated by studies using the dual-task paradigm (e.g., Madden, 1986; Salthouse, 1987; Tun, Wingfield, & Stine, 1991; Kramer & Larish, 1996; Tsang, 1998), episodic memory (Kliegl & Lindenberger, 1993; Kliegl et al., 1994; Wingfield, Lindfield, & Kahana, 1998), working memory (e.g., Oberauer, 2001; Oberauer & Kliegl, 2001; Oberauer, Wendland, & Kliegl, 2003), task switching (Kramer et al., 1999; Meiran, Gotler, & Perlman, 2001), especially as measured in 'global switch costs' (e.g., Kray & Lindenberger, 2000; Mayr, 2001; Mayr & Liebscher, 2001), and tasks where novel information is important (Cornelius, 1984; Kirasic, 1991; Willis, 1985). Age effects are larger in memory search with a variable mapping than in memory search with a consistent mapping (Fisk, Rogers, & Giambra, 1990; Fisk & Rogers, 1991). It has also been reasoned that old adults are less efficient in actively inhibiting irrelevant information in working memory (Kane, Hasher, Stoltzfus, Zacks, & Connelly, 1994; Hasher, Stoltzfus, Zacks, & Rypma, 1991; Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999). These effects will be briefly discussed in the following paragraphs.

Age effects in dual-task costs can sometimes be explained by age effects in the component tasks (Salthouse & Meinz, 1995). However if the attentional demands become sufficiently high, this no longer seems to be the case (Tsang, 1998).

Episodic retrieval seems to be more affected by aging than semantic retrieval especially under conditions of interference (e.g., Riby, Perfect, & Stollery, 2004). A characteristic feature of episodic memory is that episodes are bound to a specific spatio-temporal context. Episodic retrieval in most cases requires more cognitive control and is less overlearned than the retrieval of semantic facts.

Working memory tasks are often measured in accuracy, not latency, thus, it is difficult to compare results between working memory and reaction time tasks. However, there is ample evidence that old adults' performance breaks down in tasks that require high amounts of working memory capacity (Mayr & Kliegl, 1993; Babcock, 1994; Oberauer, 2001). What are the mechanisms responsible for this breakdown? Oberauer (2001) investigated age differences in the effectiveness of removing irrelevant information from working memory with a modified version of the Sternberg task that required memorization of two simultaneously presented lists, with a color cue later indicating which list was relevant for the recognition task. It was reasoned that items from both lists produce automatically interpreted familiarity signals, while items from the relevant list can only be identified by an effortful, capacity-limited recollection process that retrieves memory episodes linking items to a spatio-temporal context (the list color). Results were interpreted within a working memory framework suggested by Cowan (1995), who distinguishes between the focus of attention and the activated part of long-term memory. No age differences in the effect of irrelevant setsize were found, which Oberauer interpreted as an age-equivalence in the effectiveness of removing information from the focus of attention. However, intrusion effects of the irrelevant list were much larger for old adults. Thus, residual activation of no-longer relevant episodic long-term memory representations seems to produce a larger amount of interference with working memory maintenance in old adults.[7]

Some aspects of task-switching seem to be relatively unaffected by aging, while other aspects produce large age effects. Task-switching refers to switching between two relatively simple tasks. The switch cost can be decomposed into a number of processes (Meiran, Chorev, & Sapir, 2000). Of particular interest in the context of aging research are 'mixing costs' or 'global selection costs' (Mayr & Liebscher, 2001), which are given by the difference between average performance on pure task blocks, in which the component tasks are performed separately, with mixed blocks, in which the tasks are interleaved, for example, with a task cue indicating which task to perform on a particular trial. Several studies have reported large age effects in global selection costs (Kray & Lindenberger, 2000; Mayr, 2001; Meiran et al., 2001). Mayr and Liebscher (2001) show that age effects in global selection costs persist for a long time even when one of the tasks is no longer relevant (which could be regarded as an example of proactive interference). In comparison, age differences in 'local switch costs', i.e., the reaction time difference between a task-repetition and a task-change in a mixed block appear to be absent or small, especially if sufficient time is allowed for preparation of the upcoming task, and if each of the two tasks involves only two-choice reactions. While they might occur at the beginning of the experiment, age differences diminish during the course of the experiment, and relatively large differences are only observed under high memory load (Kramer et al., 1999). The size of the age difference in global selection costs is contingent on the degree of overlap of S-R rules for the two tasks on the response end. The age effect in set selection cost under conditions of full overlap is much larger than under conceptual overlap, where the same concepts (left/right), but different effectors (different hands) are used in the two tasks (Mayr, 2001). These results seem to indicate that a process involved in updating an internal control setting in the face of actual or potential interference between sets might be responsible for global selection costs. Specifically, updating and maintenance of a clear representation of the currently relevant set might be disturbed by rules that share features with the currently relevant rule. The series of studies I will report below investigates a similar hypothesis in a different context. Because the context does not require task-switching, updating of the set is not required. Instead, maintaining a mental set under conditions of stimulus-elicited conceptual interference is at the focus of interest.

Taken together, underestimation of age effects by a typical Brinley slope is often observed in tasks that rely to a relatively large degree on controlled attention or executive control processes. What all of these tasks seem to have in common is the fact that arbitrary task rules play a major role in determining task performance. At a very general level of description, arbitrary task rules are associations between elements that are only relevant in the current experimental context. These episodically associated elements do not have to use the same representational (e.g., phonological or spatial) code, instead, the association often leads to a short-term binding of different codes. Age effects in tasks with a high degree of arbitrariness are particularly severe if interference from concurrently activated rules is likely, for example because different rules share some features.

---

[7] Oberauer and Kliegl (2001) further explored the nature of the effect of aging on working memory capacity limits by comparing several formal models of working memory capacity limits. They found that two candidate mechanisms, one based on interference between elements due to feature overlap and the other on decay combined with a serial reactivation process, were the most likely sources of capacity limits. Other models, based on the distribution of an unspecific resource or on crosstalk at retrieval, fit the experimental data less well. The interference-based model gave a particularly good fit while at the same time being more parsimonious than the decay-based model. Decay and interference are the two mechanisms that are the main source of forgetting in general theories of memory.

Arbitrary task rules as a common factor in general slowing and deviations thereof

Age-related decline in the reliability of mental sets, an episodic buffer-type of executive storage system

The core proposal of this thesis rests on the observation that both the underestimations and the overestimations of age effects by the typical 'general slowing' pattern may be explained by a common factor, namely the extent to which arbitrary, episodically instructed information is critical for performance in the task. These task rules are only required in the current task context, hence they cannot be retrieved from semantic long-term memory. Instead, coding and maintenance of the task set requires some executive resources. A combination of working memory and episodic memory may play a role in the maintenance and retrieval of task rules. The fact that reaction time tasks rely to varying degrees on short-term and working memory is often ignored or at least not explicitly discussed. However, the selective review of the over- and underestimations of age effects by a standard Brinley slope supports the idea that storage and maintenance of arbitrary task elements plays a larger role in tasks in which age-effects are higher than average, and a smaller role in tasks for which smaller-than average Brinley slopes were reported.[8]

In my opinion, it is quite likely that young adults' reaction time, i.e., the operational definition of task complexity in the Brinley plot, covaries with the extent to which arbitrary, non-automatic task demands play a role. While tasks can be conceived that lead to long response times, but that do not rely on arbitrary rules and hence are not executively demanding (e.g., sequentially complex tasks, Mayr & Kliegl, 1993), those tasks are not normally at the focus of interest of the cognitive psychologist. It is a more common approach to try to selectively affect one or a number of different component processes instead of serially chaining a number of identical component processes. If the manipulated component is an early, low-level, 'automatic' component, reaction time effects are typically small. If the component is a late, 'capacity-demanding' component, reaction time effects are large, and the process likely involves some executive, 'top-down' control. Data available for meta-analysis will be a mixture of points originating from fast, low-control, and slow, high-control conditions.

Standard slowing factor in tasks with an average degree of arbitrary rules. In typical reaction time tasks, like the ones summarized in Cerella et al. (1980) which gave rise to an estimate of the slowing factor of about 1.5, arbitrary task components are often hidden in the stimulus-response (S-R) associations that are demanded by the task instructions. It is probably safe to assume that—for convenience or other reasons—most reaction time tasks use a manual set of responses. In these experiments, all stimuli that are not inherently spatial must be associated with a spatial response (e.g. red → left key; green → right key), and a cognitive 'translation' from stimulus meaning to response has to be performed. To the very least, the S-R translation mechanism needs to establish a context-specific binding of features of the concept that is activated by the input (e.g. yellow color) to features that are part of the output (e.g. left location). Furthermore, this mapping has to be maintained during the course of the experiment. Maintenance is critical not only because of omnipresent decay, but also because there is potential for interference: presumably both the 'input' concepts and the 'output' concepts share several features (e.g. yellow and blue are both colors, and left and right are both directions), so that activation of one rule might be able to cause co-activation of another rule by spreading of activation. This is particularly obvious in task-switching studies, where different S-R rule sets that converge on a common response set can change from trial to trial. However, it is also relevant in easier tasks that use only a single set of S-R rules. Here interference can arise from other members of the same set of rules, e.g. by residual activation of S-R associations that were used on previous trials, by irrelevant attributes of the stimulus, by sequencing demands and/or other concurrently relevant task rules, etc. In short, in most reaction time tasks the correct rule has to be retrieved and applied in the presence of decay and interference from other rules, and thus even apparently undemanding choice reaction time tasks rely to some degree on executive processes and episodic representations. In the following paragraphs, the core argument will be developed by showing how age effects co-vary with arbitrariness of tasks rules in different sets of tasks. Because the focus in this thesis will be on relatively standard choice reaction time tasks, the memory demands in these tasks will be elaborated on in the following paragraphs.

Overestimation of age effects in episodically undemanding tasks. While even apparently undemanding choice reaction time tasks often rely on arbitrary rules, the response in semantic tasks—in which age effects are small or absent—often does not involve an arbitrary component. For example, in semantic production tasks, the words are merely spoken. Similarly, in simple arithmetic tasks the answer is either given verbally, or it is eliminated by psychophysical methods such as time-accuracy functions. It is interesting to note in this context that age differences in semantic priming effects de-

---

[8] In the following sections, the concepts of executive control, working memory, and controlled attention will sometimes be used almost interchangeably. This is not meant to indicate that the concepts are congruent. Rather, the aspect of the concepts that is of interest in the current research is shared among all three processes. Thus, when talking about working memory demands of a reaction time tasks, I do not refer to working memory storage in the 'slave systems', but rather to executive control of working memory (which might have its own store, termed 'episodic buffer'). Executive control of working memory is considered a function of controlled attention.

pend on the response modality, i.e. they are larger with manual responding (word → left key; nonword → right key) than with simple pronunciation of the word (Madden et al., 1993). Manual responding adds at least some degree of arbitrariness to the task.

A similar interpretation also applies to visual search tasks, where an arbitrarily-instructed reaction component only contributes to the intercept of the search function. That is, the search slope—the central dependent measure—is independent of the duration of response selection processes. Only the specification of the target introduces an arbitrary, 'episodic' element, however, its representation does likely not require much capacity because one and the same target is usually relevant for a whole block or even session. While this is true for feature search, in the case of conjunction search values on two dimensions have to be associated, a fact that could explain the small, but reliable age differences in conjunction search (see Humphrey & Kramer, 1997).

The fact that slowing is not observed for stimulus evaluation, as indicated by P300 latency, in a task in which slowing is observed in reaction time, is consistent with the hypothesis that the slowing effect is contingent on response selection among a set of arbitrarily mapped response alternatives.

Underestimation of age effects in tasks with a high degree of episodic demands. In coordinatively complex (and other working memory) tasks, arbitrary, task-relevant instructions change often and partly even have to be generated during the course of a trial. This could be the reason why age effects in coordinatively complex tasks are particularly large. The same reasoning applies to episodic memory tasks proper, for which large age effects are typically observed. Similarly, studies in which large age effects in global switching costs were observed employed an arbitrary mapping of stimuli to responses, and furthermore, the stimuli from different task sets were mapped onto one and the same response set. Hence due to feature overlap there was a high potential for interference by the irrelevant set of task rules during retrieval of the relevant set. As Mayr (2001) showed, if this potential for interference at the response end is eliminated, then the age effect is much reduced.

Dominant aging theories regard cognitive slowing, as indexed by performance in a large number of reaction time tasks, as an indicator of a basal, unspecific decay process that is at the heart of age differences of more complex cognitive processes like episodic memory (e.g., Verhaeghen et al., 1997). The observations summarized here suggest the theoretically interesting possibility that 'general slowing' is a by-product of a relatively specific process, namely the progressively less reliable representation of episodic (in the sense of context dependent), task-relevant information. A moderating factor for the size of age differences would thus be the extent to which episodic components are critical in a given task. Typical reaction time tasks with arbitrary S-R associations might take an intermediate position between semantic tasks with very low episodic demands one the one hand, and episodic, coordinatively complex, or executive tasks on the other hand (for a similar argument, see Jordan & Rabbitt, 1977).

Episodic memory demands of choice reaction time tasks

Without a natural relationship between stimulus and response sets—i.e., when there is no environmentally determined or pre-experimentally overlearned set of associations—choice reaction time tasks require the translation of stimulus codes (e.g., color) into response codes (e.g., spatial codes corresponding to key locations). In this section it will be argued that the translation process (a) has many similarities to cued recall in episodic memory retrieval, and (b) is controlled by executive working memory processes. Although it is often not explicitly stated, most researchers would probably agree that the translation of stimuli to responses requires working memory capacity. Implicitly, this is reflected in the limited number of response alternatives that are usually implemented in a given task.

Do episodic task demands influence choice reaction time? Results from a long tradition of research on stimulus-response compatibility (SRC) (Fitts & Seeger, 1953; Fitts & Deininger, 1954; Proctor & Reeve, 1990; Lien & Proctor, 2002) suggest so. Before I present a selective review of these results, let me briefly introduce the concept of stimulus-response compatibility and its relation to 'episodic' task aspects. "SRC effects are differences in reaction time and accuracy, as a function of the mapping between stimulus and response sets or the members within the sets" (Lien & Proctor, 2002, p. 213). There is a distinction between *set-level* and *element-level* compatibility effects. The latter occur when the S-R sets are held constant, but the mapping of individual stimuli to responses is varied (e.g., direct, mirrored, arbitrary). On the other hand, a change of either stimulus or response set implies that both set-level and element-level compatibility is varied.

SRC effects arise primarily from a stage of information processing that is referred to as the translation stage or the response-selection stage. In the most prominent model of SRC effects (Kornblum, Hasbroucq, & Osman, 1990) there are two routes leading from stimulus representations to response-selection, namely automatic activation and intentional translation. Activation of response codes along the automatic route is fast, and even if automatic activation does not necessarily lead to suprathreshold activation and hence to a response, stimuli that are highly compatible with a response set member inevitably cause some response set priming. The automatic activation route can only be used in the case of stimulus and response sets which share features, i.e. sets that are compatible at the set-level. A classic example would be spatially arranged stimuli and spatially arranged responses.[9] With overlapping S-R en-

---

[9] However, even the internal representation of natural numbers (undoubtedly due to learning history) seems to have a spatial component, as indicated by the SNARC effect (Dehaene, Bossini, & Giraux, 1993; Lammertyn, Fias, & Lauwereyns, 2002).

Controlled attention and working memory, executive control, and episodic buffers

The concept of coordinative complexity was developed to test a function of working memory. Working memory (WM) is a high-level construct in cognitive psychology that largely overlaps (and is sometimes seen as congruent) with other concepts such as controlled attention (e.g. Engle, 2002; Cowan, 1995; for an overview see Miyake & Shah, 1999). The most basic definitions regard WM as a system for "temporary maintenance and manipulation of information" (e.g. Baddeley & Hitch, 1974; Baddeley, 2001, p.849) or the "mechanism underlying the maintenance of task-relevant information during performance in a cognitive task". A similar definition is given by Cowan (1999), who conceives of WM as "cognitive processes that retain information in an unusually active state". Psychometric tests of WM span often focus on simultaneous processing and storage requirements. According to Daneman and Carpenter (1980, 1983), results from reading span tasks lead to the conception of WM capacity as a limited resource that can be allocated to processing functions, storage functions, or both. WM capacity is tied to the specific processing demands of the concurrent task.

Turner and Engle (1989) proposed an alternative view of WM capacity as more general, reflecting an abiding, domain-free capability that is independent of any one processing task. Indeed the specific concurrent-processing task has little impact on the predictive validity of WM span measures across a number of higher cognitive abilities. Rather, the span tasks seem to tap a general cognitive primitive. Engle, Tuholski, Laughlin, and Conway (1999) argue that WM span tests 'work' because they reflect a general controlled-attention capability. In this view, WM is a hierarchically organized system, in which short-term memory storage components subserve a domain-free, limited-capacity controlled attention—the processing aspect of WM is considered relatively independent of the code used for storage.

A closer look at current theoretical conceptions of working memory as well as of the neurophysiological literature indicates that executive functions of working memory and the controlled scheduling of 'attentional resources' might be the same, or at least intricately related. Randy Engle has wrapped this in the short formula *working memory* = *short term memory* + *controlled attention* (Engle, 2002; Kane & Engle, 2003), which might however be somewhat of an oversimplification (Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; Oberauer, Süß, Wilhelm, & Wittmann, 2003). Baddeley, whose research was more concerned with spatial and especially with phonological working memory *storage* systems than with executive functions, nevertheless integrated the latter in his working memory model, and was recently even motivated to add an executive storage module (Baddeley, 2000). Importantly, the representational code in this 'episodic buffer' store is multidimensional, i.e., it can combine spatial and verbal codes. In several recent historical reviews of his model, Baddeley (2001, p.94, see also Baddeley, 2000, 2003) states that the central executive was initially conceived of

> in the vaguest possible terms as a limited capacity pool of general processing resources. For the first decade, it served principally as a convenient ragbag into which could be thrust such awkward questions as what determined when the sketchpad or phonological loop was used and how they were combined. Implicitly, the central executive functioned as a homunculus.

The main reason for this lack of clarity concerning executive control was the fact that no well-developed theory of controlled attention existed at the time Baddeley's WM framework was formulated. The functions of the central executive were inspired by a taxonomy of controlled attention given by Shallice and colleagues (Norman & Shallice, 1986; Shallice & Burgess, 1993), which includes a Supervisory Attentional System (SAS). The SAS is conceived of as a conscious control mechanism that resolves interference between activated action schemas. In particular, when a prepotent action is environmentally triggered but conflicts with the individual's goal state, the SAS biases the action-selection process by providing additional activation to a more appropriate action schema and by inhibiting the activation of the inappropriate schema. Thus the need for active maintenance of task goals, which may be minimal in many contexts, is critical under conditions of interference. Interference slows and impairs memory retrieval, enhancing the importance of keeping task-relevant information highly active and easily accessible.

This leads to the prediction that individual differences in WM capacity will be most important to higher-order cognition in the face of interference. On average, old adults' WM span is smaller than young adults' (Lustig, May, & Hasher, 2001). This has been taken to indicate that an active processing, executive, controlled attention component of WM is impaired in old age, rather than a passive storage system. In general, an executive deficit would manifest itself in tasks requiring flexibility in allocation of resources, especially the deactivation of an established task-set. Recall the most consistent result from cognitive aging research: age-related slowing correlates with (coordinative) task complexity. What constitutes task complexity? As a working definition let us assume that a more complex task either involves more cognitive processes, or a higher chance of interference between the same number of processes. Given this definition, task complexity seems rather congruent with degree of involvement of executive control. Recent results from functional neuroimaging (see Duncan & Owen, 2000 for a review) indicate that the more complex a task is, the more likely becomes the participation of frontal areas. To be more precise, a fairly small prefrontal area seems to be sensitive to manipulations of working memory load, response conflict, and even perceptual difficulty (e.g., Duncan & Owen, 2000; Miller & Cohen, 2001). The same (mostly dorsolateral prefrontal) brain systems seem to show enhanced activity in tasks that were traditionally thought to reflect rather different functions. On the other hand, the location within the brain of working memory storage systems as well as of the actual sites at which e.g. the selection function of attention operate are not prefrontal, but in 'lower' systems, like semantic storage systems with designated codes. However, whereas the *storage* areas for verbal and spatial material are different, the *processing* function for both seems to be located in a similar area, the prefrontal cortex, which is able to handle multidimensional codes. Coincidentally, this same system is also critically involved in representing arbitrary rule-like associations (e.g., Miller & Cohen, 2001) and in episodic retrieval (e.g., Cabeza & Nyberg, 2000).

In conclusion, attentional control and executive processing (not: storage) aspects of working memory as well as of episodic retrieval might be similar or identical (Kane & Engle, 2003; Conway & Engle, 1996; Engle, 2002; Nyberg, Forkstam, Petersson, Cabeza, & Ingvar, 2002), and play a major role in determining the complexity effect. It seems quite possible that old adults' impairment in an executive, prefrontal system—which may in turn be caused by a deficit in a more basal system (e.g., the dopaminergic system) that the prefrontal system critically relies on (Braver et al., 2001; Li et al., 2001)—leads to the observed pattern of complexity-proportional slowing (West, 1996). Presumably, the more coordinatively complex a task is, the more executive, prefrontal resources are required.

sembles, activation can flow along both routes in parallel. If the element-level mapping rules are consistent with the set-level relationship (i.e., each stimulus element is mapped onto the pre-experimentally associated response element), then responding can be based entirely on the activation due to response priming along the automatic route. The intentional route here thus only has a verification function. In the terminology used here, episodic task demands are minimal.[10]

In other situations with overlapping S-R ensembles, the intentional route comes into play if instructions (or task rules) require some modification of the element-level mapping—for example, if overt responses are only to be given to a subset of stimuli, or if an incompatible mapping is used (e.g., respond with the key opposite the stimulus location). Here the intentional route, in addition to a verification function, also has an inhibitory function to ensure that the instructed stimulus dimension instead of automatic activation guides the response. Arbitrary task demands are larger in these situations. Whether or not episodic retrieval of element-level rules is necessary depends on the type of mapping rule.[11]

With overlapping S-R ensembles, but very complex mapping rules, the episodic retrieval demands are similar to situations with nonoverlapping ensembles, in which no compatible S-R mappings exist. Here, only

---

[10] There might still be some episodic aspects if compatible and incompatible mappings are used within an experiment. Episodic demands will be truly minimal after a fair amount of practice with compatible mappings within an experimental setting.

[11] In principle, rule complexity and hence the episodic demands could be expressed by information-theoretic metrics like Minimum Description Length (MDL).

the controlled, intentional translation route can be taken. The controlled route is much slower, and in the case of arbitrary relations between stimulus and response sets (e.g. colors and keys), it functions by a process corresponding to episodic cued recall, termed "table lookup" by Kornblum et al. (1990). In task-switching situations with arbitrary task-cues, episodic retrieval is necessary not only at the element-level, but also at the set-level.

Do episodic task demands contribute to SRC effects? Do they require controlled attention and working memory capacity? Several findings suggesting a positive answer to both questions will be summarized below. According to the view proposed here, episodic demands and executive working memory processes are a major determinant of task complexity, which is the main factor determining reaction time. The main problem with task complexity is that it is very hard to define, as was briefly sketched above.

Early attempts at defining task complexity arose from an information theoretical framework and tried to explain reaction time exclusively by stimulus information. Arguably the most prominent example is the Hick-Hyman law, which states that

$$rt = a + b\ log_2 H$$

Reaction time (rt) is a log-linear function of stimulus information transmitted (H), measured in bits. Hick (1952) showed this for an increase in the number of (compatibly mapped) stimulus-response alternatives, and Hyman (1953) showed that stimulus information, not number, is the determining factor. He unconfounded stimulus information and number of alternatives by varying probability of occurrence of the stimulus alternatives.

However, 'noninformational', cognitive factors have subsequently been found to influence the slope of the Hick-Hyman law. One of the strongest noninformational influences is the type of mapping between stimuli and responses. Compatibility interacts with the gain factor b: the lower the compatibility, the greater the effect of increasing the number of alternatives. For example, the Hick-Hyman slope b is much larger with arbitrary than with compatible mappings (Alluisi, Strain, & Thurmond, 1964; Teichner & Krebs, 1974). Because stimuli and responses remain the same regardless of whether the mapping is compatible, inverse, or arbitrary, type of mapping does not change the amount of stimulus information. Since the external informational demands remain the same, this can be interpreted to indicate that mapping manipulations change the amount of 'internal' information. Since 'internal information' needs to be internally stored, it can be argued that memory is the critical factor leading to the increased slopes. This is supported by evidence showing that spatial S-R compatibility interacts with memory load (Crowder, 1967; Logan, 1979, 1980; see however Egeth, 1977).

Practice is another noninformational factor influencing the Hick-Hyman gain factor: the greater the amount of practice, the less the effect of increasing the number of alternatives (Teichner & Krebs, 1974). Furthermore, the amount of improvement in response selection with practice is greater for arbitrary (and for incompatible) than for compatible mappings. Even with an arbitrary S-R mapping, RT can become relatively short after extended practice (Teichner & Krebs, 1974). Only after sufficient practice within the task context will the association be firm enough so that task performance (as far as task rules are concerned) solely relies on long-term memory—i.e., a direct path from stimulus to response has been established. Both the power law of practice (e.g., Newell & Rosenbloom, 1981) and studies focusing on the development of automaticity suggest that it is unlikely that the stage at which attentional capacity is no longer needed is reached within a single experiment comprising one or two sessions. SRC effects are still evident after considerable practice. An impressive demonstration of the capacity demands of arbitrary mappings was given by Logan (1979), who used a concurrent-memory-load technique to evaluate the attentional demands of arbitrary stimulus-response mappings (with sets of 2, 4, or 8 letters mapped onto buttons). He found that the effects of memory load were the greater, the larger the number of S-R alternatives was early in practice. This interaction diminished over time, but only after five full days of practice with consistent (but arbitrary) mappings, the effects of memory load and number of alternatives were additive. Logan (1980) provided further evidence that S-R mapping rules are held in short-term memory to enable performance on reaction time tasks.[12]

Interestingly, at least one recent set of findings suggests that robust, 'direct' S-R associations quickly emerge in long-term memory (Hommel, 1998; Hommel & Eglau, 2002). Loading working memory, either with digits or with additional mapping rules, did not affect backward compatibility effects of a secondary task (say 'red' in response to the letter H) on a primary task (press left key in response to the color red) under dual task conditions. Hommel and Eglau explicitly speak of "automatic, capacity-free translations of stimulus features into arbitrarily mapped responses". Thus under appropriate conditions, an arbitrarily mapped response can be retrieved from episodic memory with relatively little need for controlled intervention, and working memory capacity does not seem to play a major role, at least for young adults.

The findings can be reconciled by noting that stimulus information is probably processed concurrently by quick, parallel, and automatic pathways and by slow,

---

[12] In the discussion of his findings, Logan (1980, p.388) makes the conjecture that "from the view of capacity theory of attention, attention and short-term memory may be the same thing; both are central in the architecture of the information-processing system, both have limited capacities that can be allocated strategically, and both have been implicated in the control of behavior. [...] The primary function of attention in reaction time tasks is to prepare and maintain a set to perform the task". From a more contemporary point of view, working memory = short term memory + controlled attention (Engle, 2002; Kane & Engle, 2003). Thus, it is likely that the effects of an arbitrary S-R mapping tap controlled attention.

serial, and controlled pathways (e.g., Logan, 1988; Lu & Proctor, 1995; Kornblum et al., 1990; DeSoto, Fabiani, Geary, & Gratton, 2001; De Jong, Liang, & Lauber, 1994; Gratton, Coles, & Donchin, 1992; Pavese & Umilta, 1998; Hommel & Eglau, 2002). The latter processes are capacity-limited and eventually lead to response selection, which is controlled by top-down attention and relies on working memory. Response selection is "an intentional act required even for highly compatible and practiced tasks and is restricted to processing one task at a time" (Lien & Proctor, 2002, p.212). While the fast, automatic processes are often thought to be limited to compatible responses, Hommel and Eglau's results indicate that even with arbitrary mappings, some degree of automaticity might be achieved relatively early in practice, at least with small sets of S-R rules. However, compatibility is a relative term, and automaticity will develop to some degree within the course of an experiment, leading to the establishment of a (weak) direct route even for stimulus-response associations that were new at the beginning of the experiment. Thus while some activation may flow along the direct route, responding with arbitrary mappings is still mainly determined by the intentional route, which draws attentional resources.[13]

Controlled processing and episodic retrieval are even more relevant when switching between sets of S-R rules. Mayr and Kliegl (2000, 2003) decomposed task-switching costs into cue-switch costs and true task-switch costs by comparing a task switch proper with a situation in which the task set repeated and only the task cue changed.[14] Considerable switch costs emerged even in the latter situation, although the set of rules that were relevant did not change. The authors interpreted the cue-switch cost as indicating that a major component of the traditional switch cost can be attributed to episodic long term memory (LTM) retrieval. If this interpretation is correct, then retrieval of a task set from LTM into working memory is necessary even when the task set does not change.

An interesting speculation is that a similar LTM retrieval effect might also be at work in a simpler context, where only a single task is relevant. In the case of arbitrary mappings between stimuli and responses (e.g. blue → left key, yellow → right key), a retrieval process might be induced by presentation of the stimulus. Although no switching of task sets is required, retrieval of the mapping rule for a given stimulus may still be obligatory. Evidence for such effects of episodic retrieval during attentional selection comes from observations of stimulus-specific repetition priming in the Eriksen flanker task (Mayr, Awh, & Laurey, 2003), as well as from episodic retrieval-based explanations of the negative priming effect (Neill, 1997; Neill & Valdes, 1992, but see Strayer & Grison, 1999). Furthermore, the slope of the Hick-Hyman law is also affected by the repetition effect. Responses to stimulus repetitions are faster than to alternations, and the size of this effect is larger with lower degrees of S-R compatibility and with larger numbers of stimulus alternatives (Kornblum, 1973).

Taken together, all of these deviations from a law relating external information to reaction time appear to be related to the degree to which the task requires reliance on internally stored information that is not over-learned. In other words, cognitive task complexity is strongly related to episodic memory processes. To conclude, it is likely that the 'table lookup' processes involved in finding the response with arbitrary S-R mappings are similar to episodic cued recall and draw capacity from the pool of attention/working memory. It is not presently clear whether these capacity demands arise during S-R translation or during response selection, but in any case they arise relatively late, after perceptual categorization of the stimulus is completed.

Episodic task demands and executive control. In the present thesis I propose a specific aging-related deficit, namely a decrease in the reliability of mental sets, consisting of episodic accumulators. The hypothesis of a reduced reliability of mental sets focuses on 'episodic' task aspects, i.e., those aspects that are not pre-experimentally established, but only relevant in the experimental context. In the Norman and Shallice's (1986) taxonomy of situations creating a demand for "supervisory", i.e. executive functions, 'episodic' task aspects correspond to "non-routine or not well-learned responding". A mental set specifies the input-output relations for a given task, thus binding reactions to stimulus representations. The specification and maintenance of parameters for a task at hand, where parameters describe the set of input-output-relations (including internal representations), has been suggested as one of the principal functions of executive control (e.g., Logan & Gordon, 2001; Braver et al., 2001; Botvinick, Braver, Barch, Carter, & Cohen, 2001). Parameter specification by executive control is only needed if stimulus and response are not naturally or pre-experimentally associated. Reliable mental sets are thus of critical importance if performance depends on the binding of arbitrary task aspects, e.g. with an arbitrary stimulus-response mapping.

The importance of reliable mental sets is enhanced in conflict situations, i.e., situations in which prepotent or 'habitual' response tendencies have to be overcome, because additional attentional capacity is required to resolve the conflict. If a limited capacity of attentional/working memory resources is assumed, then in the case of response conflict less capacity should be

---

[13] Further support for the assumption that performance in tasks using an arbitrary S-R mapping relies on working memory comes from correlational studies showing that choice reaction time with arbitrarily mapped responses is a surprisingly good predictor of fluid intelligence, while simple reaction time is not (Süß, personal communication, Dec. 2003). The evidence linking this to working memory is rather indirect however, and further analyses have to reveal whether choice reaction time and working memory capacity, which is a well-known predictor of fluid intelligence scores, share variance—as Logan's (1979) results would indicate.

[14] This was achieved by using a 4:2 mapping between cues and task.

available to other capacity-demanding processes, such as maintenance of arbitrary mapping rules. Current neurophysiologically motivated theoretical models of attentional control explicitly recognize this fact (Braver et al., 2001). For example, the influential model proposed by Botvinick and colleagues (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Botvinick et al., 2001) assumes that the anterior cingulate cortex (ACC) has the function to detect response conflict. Whenever a conflict is detected, its occurrence is signaled to the dorsolateral prefrontal cortex (DLPFC), which is responsible for scheduling attentional resources. A conflict signal fed into DLPFC from ACC indicates that more resources should be diverted to the representation of the task-relevant stimulus dimension. The passive role of the ACC has recently been questioned by Mayr et al. (2003), who showed that the empirical pattern of conflict-adaptation effects[15] on which the model is based is more compatible with a memory-based priming account than with a model that assumes a passive conflict signal, independent of the specific stimulus history. Mayr et al. (2003) found that faster responding on the subsequent trial was limited to exact repetitions of the target+flanker ensemble.[16] Cognitive control can be completely bypassed in trials featuring exact repetitions of target-distractor (and response) ensembles because of memory-based priming. Episodic priming might therefore provide a short-cut to response selection, so that less conflict occurs on repetition trials. If episodic priming is unavailable because the stimulus-response ensemble changed, then slow, controlled retrieval is necessary, allowing for a new locus of conflict at the level of memory access.

What kind of code does the system implementing arbitrary mappings use? An arbitrary mapping between stimuli and responses consists of the ‘episodic’ short-term binding of stimulus features to response buffers. Working memory seems to be involved in the maintenance of this binding and/or the selective activation of the matching rule based on its input on each trial (e.g., Logan, 1979, 1980). The process implementing arbitrary rules must be flexible enough to encode and store associations between any task elements, regardless of their internal code—thus a multimodal representation is a requirement. Baddeley (2000) recently modified his model of working memory (Baddeley & Hitch, 1974; Baddeley, 1986, 1992a, 1992b) to include a storage component of the central executive, which had previously been conceived of as a module devoted exclusively to processing, not storage. The new component, termed ‘episodic buffer’, “comprises a limited capacity system that provides temporary storage of information held in a multimodal code, which is capable of binding information from the subsidiary systems [i.e. phonological loop and visuo-spatial sketchpad], and from long-term memory, into a unitary episodic representation” (Baddeley, 2000, p. 417). Thus Baddeley’s episodic buffer shares many of the properties that are required for the instantiation and maintenance of arbitrary mapping rules.[17] If a set of recent associations is to be kept in a highly accessible state to achieve good task performance, then some sort of short-term memory store with a multimodal representational code is necessary for the maintenance of such arbitrary associations.

I conjecture that the representation in episodic buffers (or mental sets) is less reliable in old age. This could account for age effects in reaction time tasks, because task complexity may be congruent with the amount of ‘episodic’ task demands, i.e. the aspects of a task that are arbitrary and only specified in the current experimental context. In the light of Logan’s results, because most experiments take place in only one or two sessions, it is likely that arbitrary task rules require executive capacity, because they do not become sufficiently automatized during the course of an experiment. This is certainly true for arbitrary S-R mappings, but it can probably be extended to other arbitrary task aspects that are only relevant during the experiment, and that require temporary formation and maintenance of associations. With very high degrees of compatibility and highly overlearned associations such as in reading aloud, the slope of the Hick-Hyman law becomes almost flat, while it is very steep on change trials where responding is based on newly acquired arbitrary stimulus-response associations. If task complexity is largely determined by episodic task aspects, then the assumption of an age-related decrease in the reliability of episodic accumulators is well-suited to explain Brinley plot regularities.

The Episodic Accumulator model

Reliability of mental sets could thus be one of the executive functions particularly affected by aging. If the general slowing pattern is produced by a less reliable representation of mental sets in old age, then the ‘standard’ old-young ratio of 1.5-1.8 could be a consequence of the ‘average’ degree of arbitrariness in typical cognitive psychological reaction time tasks (caused e.g. by the arbitrary mapping of colors to response keys). To more specifically characterize the predictions derivable from this hypothesis, the Episodic Accumulator model was developed. The main assumptions of the model are described below, and a block-and-arrow flowchart version of the model is given in Figure 1.

1. Early stages of processing (e.g. perceptual classification and semantic categorization) are not or at least not much affected by the aging process. As

---

[15] Responses in a conflict trial following a conflict trial are faster than in a conflict trial following a nonconflict trial, and responses in a nonconflict trial following a conflict trial are slower than in a nonconflict trial following a nonconflict trial.

[16] The Botvinick et al. conflict adaptation model would assume that conflict on a given trial signals enhanced control demands, and enhanced control leads to a faster conflict resolution on the subsequent trial.

[17] There is also a similarity to the long term working memory concept introduced by Ericsson and Kintsch (1995).

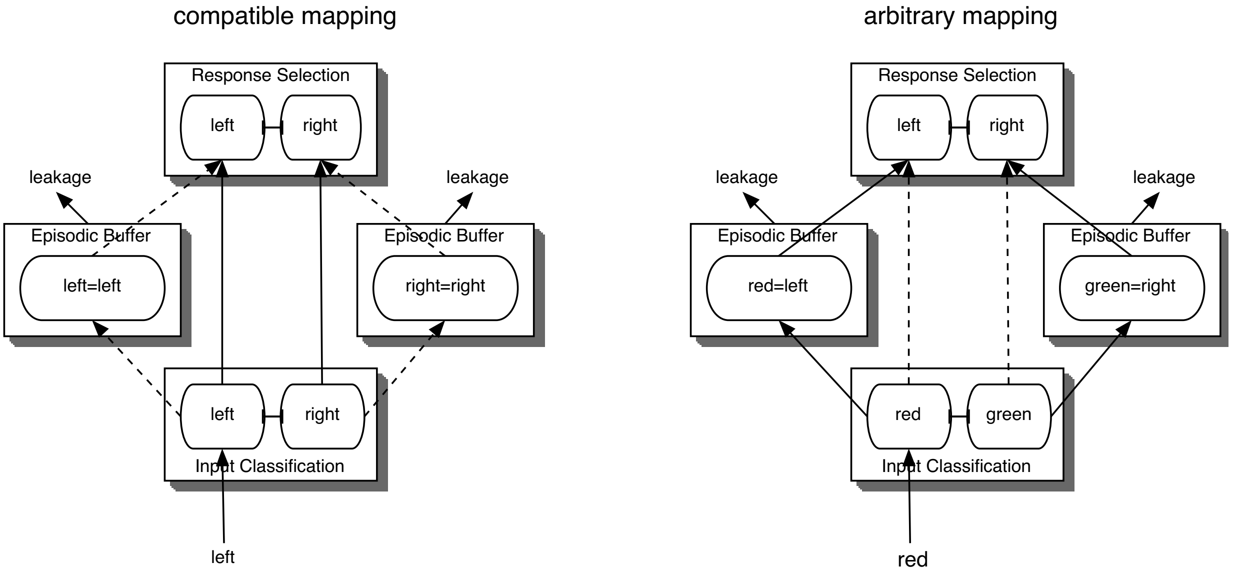


Figure 1. The Episodic Accumulator Model in tasks with compatible (left) and arbitrary (right) mappings. Solid lines indicate strong pathways, and broken lines indicate weak pathways. The arrows labeled 'leakage' indicate that some information might get lost due to unreliable accumulators. Older adults' episodic buffers are assumed to be less reliable.

has been argued above, age differences are absent in 'pure' measures of early processing (e.g., Verhaeghen & De Meersman, 1998; Mayr & Kliegl, 2000) and in the P300 (Bashore et al., 1989). Thus for each stimulus condition i (e.g. levels of perceptual difficulty) and time step t, early stages produce an evidence count E_i, which does not differ between age groups.

2. I distinguish between two situations with regard to early difficulty manipulations. This distinction is motivated by results from the additive factors research program (Sternberg, 1969), which indicate that factors affecting early, *perceptual* classification and late, response translation and selection stages rarely interact. On the other hand, factors affecting the early stages of *cognitive* processing, such as probability or expectancy manipulations, often give rise to interactions with response selection difficulty (Keele, 1969; Miller & Anbar, 1981; Rabbitt, 1967; Sternberg, 1969; Sanders, 1995; for a review see Wickens, 1992).
First, the system can be configured in a way that perceptual classification and response translation and selection are independent. In this case, early stages of processing are thresholded, such that they are only read out by translational and response selection systems after the threshold has been crossed. In this case, early systems just convey identity information, but not stimulus strength information, i.e. they act like a binary gate that signals stimulus presence. Early difficulty manipulations will lead to a delay in reaching the threshold, but after the threshold has been reached, the same information will be available for later systems in both perceptually easy and perceptually difficult conditions.
Second, another configuration is relevant for situations in which the early difficulty manipulation affects a 'cognitive' variable (as opposed to a perceptual variable). In this case, information from (not quite so) early systems can be read out immediately by the response system, i.e. processing is cascaded (McClelland, 1979), and the early cognitive stage provides a strength signal.
There is considerable evidence that behavior in cognitive tasks can shift between exhibiting a staged and a cascaded mode of processing, depending on task characteristics such as stimulus category or speed pressure (Miller & Hackley, 1992; Kello, Plaut, & MacWhinney, 2000). Similarly, processing models of an identical architecture can operate in staged or cascaded mode, depending on the gain of the S-shaped activation function in a neural network (e.g., Kello et al., 2000; Schwarz, 2003). In the thresholded case the net input may be so large that the activation function is already near asymptote, so that small changes in input do not make much of a difference in activation. In the cascaded case, the net input is in the 'sensitive region' of the activation function, where activation is approximately linearly related to input.

3. Information originating from early systems is transformed into response-relevant signals by response accumulators. Response accumulators are units that continuously collect evidence from early stages, until a fixed response threshold K is reached. Each response alternative is associated with a separate accumulator, and each accumulator receives input in parallel from two sources, namely the stimulus classification layer (see 1. and 2., above), and the episodic accumulator layer (see 4., below). For example, in a binary choice reaction time task, there are two response accumulators, one collecting evidence for the left response key, and the other for the right key. The assumption of a fixed

threshold across conditions and age groups is motivated by physiological results (Hanes & Schall, 1996).

The physiological module implementing the response accumulators might be the basal ganglia, which have been proposed to be the locus of the final pathway of response selection (Gurney, Prescott, & Redgrave, 2001b, 2001a).

4. Under conditions of arbitrary mappings, a 'translational' module has to mediate between stimulus information and response accumulators. We call this translational module 'episodic accumulators', or 'episodic buffers'. Like with response accumulators, each element-level stimulus-response rule is associated with a separate episodic accumulator. Episodic accumulators channel information from the associated concept at the semantic classification layer to the associated response accumulator. Importantly, episodic accumulators are considered to be flexibly programmable according to task demands. If they were not, we would not be able to perform a wide range of tasks given only brief instructions.

I consider episodic accumulators to be located in prefrontal cortex (PFC), most likely in dorsolateral PFC (DLPFC), which projects to the basal ganglia. Neurons in PFC have the desirable property that they are programmable according to task instructions and can hence code arbitrary rules (e.g., Miller & Cohen, 2001). Behavioral (Asaad, Rainer, & Miller, 1998; Quintana & Fuster, 1992; Bichot, Schall, & Thompson, 1996; Barone & Joseph, 1989; White & Wise, 1999; Wallis, Anderson, & Miller, 2001) as well as clinical evidence (Wise & Murray, 2000; Petrides, 1982, 1985) suggests that the PFC, which is the cortical structure showing the earliest signs of age-related biological decline, is critical for learning rules. PFC neural activity represents the rules or mappings required to perform a task, and not just single stimuli or forthcoming actions. Miller and Cohen (2001, p.178) assume that "activity within the PFC establishes these mappings by biasing competition in other parts of the brain responsible for actually performing the task. These signals favor task-relevant sensory inputs (attention), memory (recall), and motor output (response selection) and thus guide activity along the pathways that connect them (conditional association)." If the PFC represents the rules of the task in its pattern of neural activity, it must maintain this activity as long as the rule is required. The capacity to support sustained activity in the face of interference is one of the distinguishing characteristics of PFC (Miller, Erickson, & Desimone, 1996). By contrast, sustained activity in extrastriate visual areas is easily disrupted by distractors (Constantinidis & Steinmetz, 1996). Thus posterior cortical neurons seem to reflect the most recent input regardless of its relevance, whereas the PFC actively maintains task-relevant information.

While stimulus identification and response selection proper are unaffected by aging, the only influence of aging in the model is the deterioration of the mental representation that is supported by episodic accumulators. The task-appropriate use of the accumulators is a problem of episodic memory. Here, we make the central assumption that episodic information is less reliably represented in old age. This has been confirmed in a number of aging studies (e.g., Kliegl et al., 1994). If the representation of information in the episodic accumulators is less reliable, then there is a greater chance that incoming evidence is 'missed'. Thus for each age group a and each response condition r we can give a probability p_{ar} that evidence from early systems is registered by the accumulator. The loss of reliability could be responsible for phenomena such as goal neglect (De Jong, Berendsen, & Cools, 1999), lapses of intention (West & Baylis, 1998; West, 1999), or interference from concurrently activated rules (e.g., via lateral inhibition between episodic accumulators).

5. For the empirical test of the model, a corollary assumption has to be made, namely that under certain circumstances, episodic accumulators can (largely) be by-passed. Formally, in these cases, either p will have a value close to 1.0 independent of age, or the weight of the 'indirect' route via episodic accumulators is negligible in comparison to the weight of the direct S-R route. For example, this could be the case if an S-R association is pre-experimentally over-learned, or if natural relationships between stimuli and responses are exploited, such as in highly compatible spatial mappings. Another such situation could occur in the case of exact between-trial repetitions of target ensemble and response (Mayr et al., 2003). This assumption is similar to the automatic route assumption in the dimensional overlap model of stimulus-response compatibility effects by Kornblum et al. (1990). It is important, because it allows to create experimental conditions which differ in the degree to which episodic accumulators are relevant for performance.

Taken together, in cascaded mode, the time T to initiate a response in a stimulus condition i, a response condition r, and an age group a is given by

$$T_{air} = \frac{K}{E_i \, p_{ar}} \quad . \tag{8}$$

Per unit time, $E_i p_{ar}$ counts of evidence are registered, thus it takes $\frac{K}{E_i \, p_{ar}}$ units of time to reach threshold. This simple model can reproduce the age-by-complexity effect, where the slope of the Brinley function is given by $1/p_{ar}$. For example, let us assume that old adults' episodic accumulator have a reliability of .5, i.e. on average they miss every other count of evidence, and that early systems produce input at a rate of $1/\Delta t$ in a difficult stimulus condition, and $2/\Delta t$ in an easy condition. Further assuming a response threshold of 600, we obtain the following response times:

$$\begin{aligned} T_{young,easy} &= \frac{600}{2} = 300 \\ T_{old,easy} &= \frac{600}{2 \times .5} = 600 \\ T_{young,difficult} &= \frac{600}{1} = 600 \\ T_{young,difficult} &= \frac{600}{1 \times .5} = 1200 \end{aligned}$$

It is apparent that the model can reproduce the age-by-complexity effect (in this case, with a slowing factor of 2.0) without having to assume a basal, unspecific deficit. In particular, it was assumed that early stages of processing function equally well in young and in old adults. However, due to the specific age-related deficit in the episodic accumulators, early stages need to produce more evidence than in the case of reliable accumulators. Unreliable accumulators therefore amplify early difficulty effects (in the case of cascaded processing).

Now consider the case of thresholded processing, where there is no continuous flow between perceptual and cognitive modules, but rather the early perceptual difficulty manipulation introduces a delay δE_i, after which E_{max} evidence counts per unit time are received by the response selection, independent of perceptual difficulty condition i:

$$T_{air} = \frac{K}{E_{max}\, p_{ar}} + \delta E_i \quad . \tag{9}$$

Cognitive processing is not affected by this delay. Thus assuming an early difficulty effect of $\delta E = 100ms$, we obtain

$$\begin{aligned} T_{young,easy} &= \frac{600}{2} = 300 \\ T_{old,easy} &= \frac{600}{2 \times .5} = 600 \\ T_{young,difficult} &= \frac{600}{2} + 100 = 400 \\ T_{young,difficult} &= \frac{600}{2 \times .5} + 100 = 700 \end{aligned}$$

In the null condition, where episodic accumulators are by-passed, we can assume that output of the early stage is directly fed into response accumulators. Thus response times are a function of a single variable, early difficulty,

$$T_{ai} = \frac{K}{E_i} \quad . \tag{10}$$

If a complete lack of perceptuo-motor slowing were assumed, then young and old adults' performance would be equal in highly natural or compatible conditions. Although the assumption of perceptuo-motor age equivalence is probably unrealistic, we use it as a first approximation, because of the small effects usually obtained in purely perceptual measures (see above).

How to test the Episodic Accumulator model

How can the model be tested? If episodic accumulators were involved to the same degree in all tasks, then testing would not be feasible. However, assumption 5 states that episodic accumulators are not needed if task performance mainly relies on automatic or pre-experimentally well-established rules (compatible conditions). Under these conditions, episodic accumulators can be bypassed, thus response accumulators can be directly activated by incoming stimulus information. A testable prediction generated from this assumption is that (a) age effects will be much smaller under compatible conditions, and (b) there will be no interaction of early difficulty manipulations and age. Thus to test the model, experimental conditions need to be created that on the one hand vary whether episodic accumulators are involved or not, and on the other hand vary 'early' difficulty, i.e. the rate of information input into episodic or response accumulators. If these conditions can be successfully created, then a characteristic triple interaction between age, early difficulty, and 'episodic difficulty' is expected, as depicted in Figure 3. However, this interaction will only appear if the episodic buffer system operates in cascaded mode, i.e. if it receives stimulus strength information as input. In the case of serial stage-like operation, where the output of earlier stages just consists of stimulus identity information, the predicted pattern of results will look like in Figure 2. In this case, episodic difficulty and age interact, but this interaction does not modulate perceptual difficulty effects.

Perceptual difficulty and arbitrary task rules

The first experiment to be reported below test the episodic buffer hypothesis by selectively varying perceptual difficulty and S-R compatibility. If a triple interaction of age with the two factors were found, amounting to an amplification of perceptual difficulty effects in the arbitrary mapping condition in old age, this would be difficult to reconcile with serial, stage-like processing of stimulus identification and response selection. However, due to the fact that manipulations that selectively influence stimulus identification and response selection stages produce additive effects in young adults under most circumstances, this result is not likely. We have modeled this by assuming that perceptual classification does not provide a strength signal, but only an identity signal to the episodic buffer stage. Under these 'serial' circumstances, a reduced reliability of episodic buffers in old age still predicts a two-way interaction of age and mapping, however, the perceptual difficulty factor is additive with respect to both age, mapping, and the interaction of the two.

Cognitive difficulty and arbitrary task rules

While it is difficult to test the 'amplification of early difficulty' prediction of the episodic buffer hypothesis with differential manipulations of perceptual difficulty and arbitrariness of S-R mapping, I will now argue

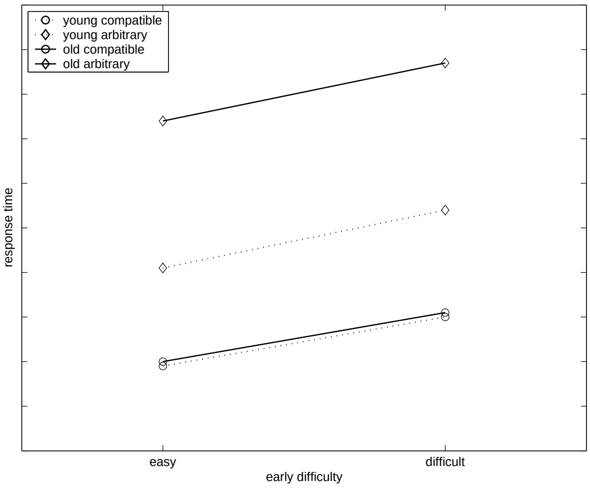


Figure 2. Predictions of the Episodic Accumulator Model in the case of serial, stage-like processing.

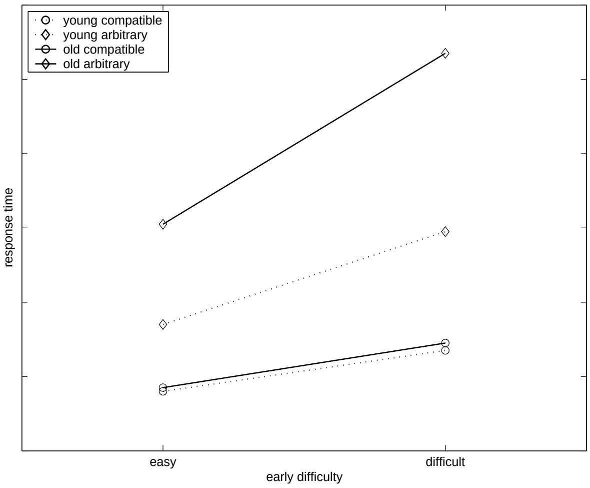


Figure 3. Predictions of the Episodic Accumulator Model in the case of cascaded processing.

that the diagnostic triple interaction pattern can be obtained by orthogonally varying early *cognitive* difficulty and arbitrariness of S-R mapping. When the episodic buffer and response selection stages start receiving input before semantic classification is fully completed (i.e. the model operates in cascaded mode), the assumption of reduced reliability of episodic buffers in old age predicts an over-additive triple interaction of perceptual difficulty, mapping and age. What kind of manipulations may lead to cascaded processing? In my opinion, the greatest chance to induce this processing mode is by making semantic classification difficult, so that classification takes rather long, thereby increasing the likelihood that information reaches episodic and response units before classification is complete. One option to prolong the time demands for successful stimulus classification would be to use near-threshold stimuli. A second option, and the one chosen here, is to use a standard tool in cognitive psychology, namely a conflict or interference paradigm (e.g., Eriksen & Eriksen, 1974; Stroop, 1935), where different stimulus features convey incongruent information about stimulus identity.

In a conflict task, there is ambiguous output of the perceptual stage due to a stimulus that has several, conflicting attributes. Typically, the stimulus has one relevant attribute (or target) and one or several irrelevant attributes (or distractor/s). It is commonly assumed that the perceptual stage performs several classifications in parallel (e.g., Treisman & Gelade, 1980). Thus if target and distractor are incongruent in a conflict task, then ambiguous output of perceptual classification presents a conflict with respect to the required response at the input side of the cognitive stage. In other words, a *cognitive* filter has to be applied to the perceptual output to resolve a response conflict. The cognitive and even the perceptual filter can be tuned by task instructions, however, filtering is not perfect. The occurrence of a conflict is particularly likely when a task-irrelevant stimulus attribute is processed in a relatively automatic fashion, while more effort is needed to process the task-relevant attribute.

To facilitate comparison of results from different experiments, the experiments reported below all used one particular interference paradigm as an early difficulty manipulation, namely the Stroop task (Stroop, 1935). Thus the argument will focus on the Stroop task, which is a paradigmatic example of a conflict paradigm. The general idea is however applicable to a broader domain of tasks. Further research with other paradigms (such as the Eriksen flanker task) will have to prove the external validity of the results. In the Stroop task, the ink color of a printed word has to be named, while word meaning is to be ignored. If the word itself designates a color different from the ink color, e.g. green printed in red, then conflict arises at the cognitive stage. This conflict has to be resolved by the cognitive system, and a decision has to be made because only a single response can be given according to task instructions. The conflict can be detected if one compares performance in this incongruent condition with a congruent condition, in which both word and color activate the same color concept.

Focusing on the Stroop task has several advantages. First, the Stroop effect is one of the most robust effects in cognitive psychology, thus there is little doubt that the ‘early difficulty’ manipulation will be effective. Second, because the Stroop task is one of the best-researched phenomena, a lot of empirical data is available. To a lesser degree, this also applies to age effects in the Stroop task. Third, relatively advanced processing models of the Stroop task exist (e.g., Cohen et al., 1990; Cohen & Huston, 1994; Phaf, van der Heijden, & Hudson, 1990; Botvinick et al., 2001; Roelofs, 2003; Gilbert & Shallice, 2002; Stafford, 2003), thus making it easier to determine the locus of the possible effects. Fourth, the Stroop task is nowadays considered to tap executive processes, in fact it is often used as a paradigm example of an executive task. Using the Stroop task as a tool to investigate age differences in a different executive function, maintenance of mental sets, might help elucidate some aspects of the rather muddled and ill-defined concept of executive control. Fifth, focusing on a single task in a series of experiments allows the construction of Brinley plots that are relatively uncontaminated by other processes. Because the Stroop interference paradigm was chosen as a cognitive difficulty manipulation in five experiments to be reported below, I will present a short overview of the Stroop task as well as relevant empirical and theoretical results in the next section.

The Stroop task

The Stroop task looks deceptively simple and yet it has generated (and continues to generate) an enormous research interest. In his review of "half a century of research on the Stroop effect", MacLeod (1991) counted over 700 relevant articles, 70% of which had been published between 1973 and 1991. In an update, MacLeod and MacDonald (2000) speak of "literally thousands of studies". Thus the current review necessarily has to be selective. I will first describe the basic phenomenon. In the following section, I will focus on the facts that (a) Stroop interference arises at or is at least promoted to the cognitive (as opposed to perceptual) stage, and thus (b) conflicting information reaches the stage where arbitrary response mappings are presumably implemented.

It has been known since the end of the 19th century that naming a color takes longer than reading a word. Estimates indicate that word reading is on average about 100-200 ms faster than color naming (Dyer, 1973; Fraisse, 1969; Glaser & Glaser, 1982), probably due to the fact that reading is very well practiced or 'automatic'. To study interference between the two processes, Stroop (1935) introduced bivalent color-word stimuli, i.e. words printed in a color. In the color-naming condition, the task consists of naming the word color while ignoring word meaning, while the reverse instructions are given in the word-reading condition. The critical manipulation affects the congruency of word meaning and color. For example, the word RED can be printed in red or in green ink, giving a *congruent* and an *incongruent* stimulus, respectively. Several neutral conditions can also be added, for example the word RED printed in black ink, or a word without color association (e.g. CAT) or a nonword (XXX) printed in green.

While the term Stroop task is reserved for the color-word interference task, a number of related, Stroop-like paradigms have been developed. These include the picture-word interference task (Rosinski, Golinkoff, & Kukish, 1975; Glaser & Düngelhoff, 1984, e.g., CAMEL printed inside of the picture of a dove), the numerosity Stroop task (Flowers, Warner, & Polansky, 1979, How many digits do you see? 5 5 5 5), or the spatial Stroop task (Shor, 1970; Logan & Zbrodoff, 1979), which typically employs a combination of direction words (such as LEFT, RIGHT) and arrows. In addition to the color-word task, only the latter is used in the current series of experiments.

Response times in the Stroop task are typically about 100 ms slower in the incongruent than in the congruent condition. This difference has been dubbed the *Stroop effect*. Inclusion of a *neutral* condition allows for a separation of the Stroop effect into *facilitation* and *interference*, designating the response time (or error) differences between congruent and neutral, and incongruent and neutral conditions, respectively. Typically, interference is much larger than facilitation if a nonword baseline is used. However, the relative size of the facilitation and interference effect depends on the type of baseline condition. For example, facilitation is larger when the neutral condition uses word instead of nonword distractors. This suggests that 'neutral' words themselves cause some interference with the color naming process. Obviously, it is unlikely that interference by neutral words without a color association originates at a semantic level–instead a lexical or phonemic locus is likely.

A verbal mediation of the Stroop effect is thus likely, but it cannot be due to verbal factors alone. For example, it is well established that Stroop interference is still obtained (though reduced) when manual, key-press responses are given (White, 1969; Logan, Zbrodoff, & Williamson, 1984; Keele, 1972). In Keele's study using manual responding, only color word distractors caused Stroop interference, while noncolor words did not. This suggests that stimulus-response compatibility plays a role in Stroop effect. If the internal codes used for distractor and response are more closely related than the codes for target and distractor, this leads to heightened interference. However, it is not clear whether manual-response Stroop tasks may involve implicit verbal coding at the response-selection stage.

Further support for an SRC contribution comes from the result that the Stroop effect is asymmetric: at least when responding vocally, colors interfere with word reading to a much lesser degree than words interfere with color naming, i.e., the *reverse Stroop effect* is smaller than the Stroop effect (e.g., Dunbar & MacLeod, 1984). This suggests that reading words is more automatic or at least has faster access to a verbal response than naming colors. This finding motivated the construction of dual-route models of the Stroop effect, which will be discussed in more detail below.

The Stroop task is nowadays often cited as a paradigmatic example of a task requiring executive control. This is because in the conflict condition, word reading due to an over-learned S-R association leads to a fast pre-activation of the word response, which has to be actively inhibited to perform in accordance with task rules. Active inhibition is considered an executive function, and has been suggested as one of the functions that are particularly susceptible to the influence of age-related decline (Hasher et al., 1991; Kane et al., 1994).

Age effects in the Stroop task

With regard to age differences in Stroop interference, there are two conflicting positions. On the one hand, the Stroop effect seems to be larger in old age. For example, MacLeod (1991, p.185) concludes that the relevant literature can be summarized by results of the first (lifespan) study that investigated Stroop effects in a developmental context (Comalli, Wapner, & Werner, 1962):

> Interference begins early in the school years, rising to its highest level around Grades 2 to 3 as reading skill develops. With continued development of reading, interference declines through the adult years until approximately age 60, at which point it begins to increase again.

The view that the increase in Stroop interference in old age is highly consistent is also stressed in the following quote from a recent study that more specifically focussed on cognitive aging, which also illustrates that Stroop interference may lead to over-proportional age-related slowing (West & Baylis, 1998, p. 206)

> Over the 60-year history of the Stroop task, only a handful of studies have addressed the relationship between increasing age, in later adulthood, and task performance. However, what this literature lacks in depth it makes up for in consistency, with all existing studies reporting an increase in Stroop interference in older adults relative to younger adults. [. . .] This age-related increase in interference does not seem to merely reflect general slowing with increasing age, as interference continues to be greater for older than younger adults when proportional interference effects, which control for age differences in baseline response latency, are considered (Spieler, Balota, & Faust, 1996). In comparison to the age-related increase in interference, facilitation appears to remain relatively stable with increasing age (Spieler et al., 1996).

Only a few studies have looked at moderator variables. The most interesting result for the current investigation was reported by Hartley (1993), who compared a fairly standard version of the color-word Stroop task (using manual responses) with a color-block version. In the latter, color and word were spatially separated, and a spatial pre-cue indicated at which location the target dimension (color) would appear. In the color-block task, attentional filtering based on spatial location is applicable.[18] In the color-word version, Hartley's results were consistent with the results cited above, namely he obtained a significant interaction of Age and Stroop condition. With the color-block task, however, there was no age difference in the Stroop effect. This indicates that age differences in the Stroop effect can be eliminated by early attentional filtering, which seems to be relatively intact in old age. The color-block condition was then used as a baseline to control for general slowing, by calculating proportions of mean reaction times in the corresponding conditions in the color-word and the color-block task. In an analysis of this measure, reflecting the relative slowing when spatial filtering is not possible, the Stroop effect was still larger for old than for young adults. Because the proportional measure controls for slowing in the baseline conditions, this seems to indicate a specific age deficit in an anterior attentional system that cannot be accounted for by general slowing. Note that manual responses were used so that some episodic memory contribution was likely, although only two colors and colored key caps were used.

A similar approach to control for baseline speed differences was used by West and Baylis (1998, Experiment 1), who compared age effects in the incongruent condition of a Stroop task to age effects in the component tasks, (a) color naming, and (b) word reading. Consistent with earlier findings (Cohn et al., 1984), and difficult to reconcile with a general slowing account, there was no age difference in word reading times, while old adults were reliably slower in color naming. However, the increase in color naming times did not suffice to account for the even larger increase in the Stroop effect. When the relative change in reaction time in the Stroop interference condition compared to color naming baseline in units of baseline reaction time was analyzed ((incongruent - color only)/color only), interference continued to be greater for older adults. Very large and over-proportional age effects in Stroop interference were also found by Brink and McDowd (1999), in a condition using manual responses in a four alternative forced-choice paradigm. Furthermore, the age difference in the Stroop effect was larger with four than with two arbitrarily mapped responses.

On the other hand, a recent meta-analysis of the Stroop effect in cognitive aging has found merely proportional slowing (Verhaeghen & De Meersman, 1998). This analysis expressed the Stroop interference effect from 20 published studies in units of mean standardized difference and found no significant age differences in this measure. Several moderator variables were also taken into account, namely , (a) baseline condition (color patches vs. XXXX or colored words), (b) presentation format (single trial, one item at a time vs. Stroop

---

[18] Spatial attentional filtering can presumably be performed by a posterior attentional system comprised of the parietal cortex, the superior colliculus, and the pulvinar nucleus of the thalamus. In contrast, in a standard color-word Stroop task in which target and distractor are integrated on the same perceptual object, a more frontal, anterior attentional system is responsible for selection between streams of processing. In this conception, the posterior system is involved regardless of whether an exogenous cue, e.g. a flash, or an endogenous cue, e.g. an arrow, is used (as long as the endogenous cue is automatically interpreted). Mueller and Rabbitt (1989) speculate that reflexive orienting of attention takes the superior colliculus pathway, whereas voluntary orienting to endogenous cues depends on posterior parietal cortex. Both of these mechanisms are part of the posterior attentional system.

test, more items at a time), and (c) presentation procedure (computerized presentation vs. printed materials). Although all of them had an influence on the Stroop effect, none resulted in a reliable Age × Condition × Moderator Variable interaction. Furthermore, the slope of the Brinley function, which was 1.88, was not modulated by Stroop condition.

However, a closer look at the results indicates that at least the Brinley results are not necessarily very robust. Verhaeghen and De Meersman used an interaction analysis approach to Brinley plot regression. Briefly, the approach tested whether the slope and intercept parameters of the linear regression functions predicted by the Cerella (1990) multilayer slowing model (see equation 6) were different between baseline and interference conditions. For this test, the regression model

$$RT_{old} = a_1 + b_1\,RT_{young} + a_2\,Cond + b_2\,Cond \times RT_{young}$$

was used, which included the dummy variable *Cond* that took a value of 0 in the baseline condition and a value of 1 in the interference condition. Therefore, a_1 and b_1 are estimates for Brinley intercept and slope, respectively, in the baseline condition, while the Brinley intercept in the interference condition is given by $a_1 + a_2$, and the slope by $b_1 + b_2$. The slope parameter was almost twice as high for the interference (2.28) than for the baseline latencies (1.19). Nevertheless, because the $a_2 = -608$ ms was not significantly different from zero[19], Verhaeghen and De Meersman felt justified to repeat the analysis without the $a_2\,Cond$ term. In this analysis, the b_2 parameter was no longer significant, so that $b_2\,Cond \times RT_{young}$ could also be dropped. The resulting model corresponded to a simple linear regression, with an estimate of $b_1 = 1.88$ corresponding to central slowing. One problem with the analysis is that it might be problematic to remove the $a_2\,Cond$ term from the analysis, because (a) the parameter was fit based on interference conditions from a relatively small sample of N=20 studies, and only marginally failed to reach significance, and (b) there is a strong correlation of *a* and *b* parameters if Brinley plots obtained in different studies are compared (Ratcliff, Spieler, & McKoon, 2000). However, the change in R squared ($R^2_{Change} = .019$; $F(2,36) = 2.10$; $p = .137$) obtained by dropping the $a_2\,Cond$ and $b_2\,Cond \times RT_{young}$ terms was rather mediocre, thus overall the impression remains that the slowing pattern obtained in the studies summarized in the meta-analysis is not specific to Stroop interference, but rather an artifact of general slowing. For the current argument, it is important to note that most of the studies used vocal responding, i.e., there were no arbitrary S-R mappings involved.

In summary, it is undisputed that young adults produce smaller Stroop interference effects than old adults. However, it is unclear whether age differences in Stroop interference merely reflect general slowing, or whether they indicate a specific deficit in some specific, possibly executive, function that is tapped by the Stroop task. If one is willing to give more weight to meta-analytic results than to single-study results, then currently the evidence seems to be pointing towards general slowing.

Sources of interference

The episodic buffer model predicts that early difficulty manipulations will be amplified by later stages that implement arbitrary mappings. However, as has been argued above, this interaction is only predicted if the early difficulty manipulation acts late enough in the stream of processing, so that stimulus classification and response selection can proceed in cascaded mode. Why do I consider Stroop conflict a suitable early difficulty manipulation? A minimum requirement for such a manipulation would be that stimulus strength information reaches response selection before stimulus identity is fully classified. Thus, if it can be shown that Stroop conflict starts before response selection, but is still present at the response selection stage, it qualifies as an early difficulty manipulation that activates the episodic buffers in cascaded mode. In this section I will review the relevant findings from the literature.

At which processing module does Stroop interference arise? This is an open research question. For example, late selection theories (e.g., Dyer, 1973) posit a response-selection bottleneck as the locus of interference, while other data and theorizing points to either a lexical or a semantic locus (e.g., Dalrymple-Alford, 1972; Luo, 1999). To quote a recent article by Logan and Zbrodoff (1998, p. 979) regarding the locus of Stroop interference:

> Some researchers have argued that it occurs early in processing, during perceptual encoding (Hock & Egeth, 1970). Others have argued for a more central locus, involving translation between codes (Glaser & Glaser, 1989; Kornblum et al., 1990; Sugg & McDonald, 1994; Treisman, 1969; Virzi & Egeth, 1985). Still others have argued for a later locus, in the response selection stage (Cohen et al., 1990; Duncan-Johnson & Kopell, 1980, 1981; Logan, 1980; Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975; Warren & Marsh, 1979).

In my view, the quote suggests that, although it is not clear at which stage the conflict originally arises, there is good evidence that information about both color and word reaches the response selection stage.

Behavioral evidence for a response-selection locus of Stroop interference comes from the *response set membership* effect (e.g., Klein, 1964; Proctor, 1978): Compared to a color distractor that is a member of the response set, there is less interference when the distracting color word does not correspond to one of the colors used as targets in the experiment. For example, when

---

[19] Statistics for the test were not reported in the original article. My replication of the analysis using the data from Verhaeghen and De Meersman (1998, Table 1, p.121) gives $t(36) = -1.99$; $p = .054$.

the target colors in an experiment are red and green, the word GREEN in red ink produces more interference than the word BLUE in red ink. The response set membership effect is obtained with both manual and vocal responding (Sharma & McKenna, 1998).

The idea that Stroop interference is present at response selection is also supported by a large number of neurophysiological findings. For example, in an experiment where they obtained standard behavioral results, Duncan-Johnson and Kopell (1980, 1981) observed no effect of Stroop condition on the latency of the evoked potential P300. This finding appears to indicate that (at least sometimes) Stroop interference originates only after visual encoding, because the P300 latency is thought to indicate completion of stimulus evaluation.

Furthermore, most neuroimaging studies of the Stroop task showed enhanced activity in the ACC on conflict as compared to neutral trials (for reviews see Bush, Luu, & Posner, 2000; Jonides, Badre, Curtis, Thompson-Schill, & Smith, 2002), and ACC is commonly associated with response selection (Picard & Strick, 1996, 2001; MacDonald III, Cohen, Stenger, & Carter, 2000; Carter et al., 1998), with different subregions serving different response modalities (Turken & Swick, 1999). A recent brain imaging study[20] lent direct support to the idea that conflicting information reaches the response-selection stage. In a spatial Stroop task (which requires manual responding to the meaning of a word while ignoring a simultaneously presented arrow), brain areas associated with correct and incorrect responses (left and right motor cortices) were simultaneously activated in conflict conditions, whereas nonconflict trials elicited brain activity only in the contralateral motor cortex (DeSoto et al., 2001). Thus most theories of the Stroop effect regard the Stroop phenomenon as a problem in the attentional control of action as opposed to selective attention of visual processing, and all current models of the Stroop effect (e.g., Cohen et al., 1990; Phaf et al., 1990; Roelofs, 2003) assume that information about both color and word is available at a post-perceptual stage.[21]

While these results clearly show that conflict is promoted up to response selection, this does not rule out that the conflict originates at an earlier level of processing. There is empirical evidence for the involvement of at least the lexical and the semantic concept levels in addition to the response selection stage (e.g., Luo, 1999; Dalrymple-Alford, 1972; Klein, 1964)[22]. An empirical phenomenon that directly supports a semantic contribution to Stroop interference is the *semantic gradient effect* (Klein, 1964; Dalrymple-Alford, 1972; Glaser & Glaser, 1989). The semantic gradient effect refers to the fact that compared to a control condition showing non-word letter strings (XXXX), there is a gradient of interference from lexical words. Color words which are members of the response set produce the highest interference, followed by color words which are not members of the response set, followed by color-related words such as FIRE or SEA, followed by color-unrelated words (e.g., THIN) which still produce some interference. Furthermore, a "semantic relevance effect" has been observed (Neumann, 1980; La Heij, van der Heijden, & Schreuder, 1985). Irrelevant color words that are not members of the response set (e.g. GOLD) cause more interference than semantically related noncolor words (e.g. FIRE). There is some evidence that the semantic gradient and semantic relevance effects may be limited to vocal responding (Sharma & McKenna, 1998).[23]

To arrive at a response when responding is manual, semantic information likely serves as input to the response maps, at least if keys are not labeled with color patches. Empirical evidence for the point that conflicting semantic information is the input to the response maps comes from research triggered by translation accounts of the Stroop phenomenon (e.g., Virzi & Egeth, 1985), which hold that colors and words are mapped from one internal code to another. They predict interference whenever the relevant stimulus type does not match the response type (e.g., a color target has to be translated into a word response, while a word target need not be translated). This explains why irrelevant words interfere with color naming, and why irrelevant colors do not interfere with word reading. Sugg and McDonald (1994) examined all four combinations of task (word or color) and response (word or color), using manual re-

---

[20] using a new method called the event-related optical signal, EROS, that has a temporal resolution similar to the ERP, and a spatial resolution similar to fMRI

[21] Of course, the boundaries between action control and perceptual filtering may be somewhat unsharp, because it has been shown that after a certain delay, the system implementing top-down attentional control can modulate the gain of perceptual channels by a feed-back mechanism (e.g., Posner & Raichle, 1994). Although there has been recent evidence for top-down attentional modulation in early cortical areas like V1 (e.g., Somers, Dale, Seiffert, & Tothell, 1999; Martínez et al., 1999) and MT (e.g., Treue & Maunsell, 1996; Treue & Martínez Trujillo, 1999), and possibly even in subcortical areas like LGN (O'Connor, Fukui, Pinsk, & Kastner, 2002), attentional affects tend to be weaker in early visual areas and more pronounced the further one looks up the visual stream. In models of the Stroop task, this top-down influence on perceptual selection can be modeled by shutting off the perceptual input of the irrelevant stimulus after a certain delay, termed "distractor duration" in the Roelofs, 2003 model.

[22] There is also evidence for interference at the phonological level, which I will ignore here for purposes of simplicity.

[23] Sharma and McKenna (1998) investigated semantic gradient and response set effects in a comparison of manual and vocal Stroop tasks. They compared responding on verbally labeled buttons with vocal responding. Sharma and McKenna replicated the semantic gradient effect for vocal responses, but found that for manual responses, the only significant component consisted the of response set membership effect, which was even larger than in the vocal condition. Thus with manual responding, interference does not seem to arise at a lexical level, but only at a 'later', response selection stage. However, this can be rephrased to indicate that Stroop interference always arises at response selection, but the location of the system where responses are selected depends on the task at hand (see Roelofs, 2003). In the vocal task, responses are selected in the lexical system, while in the manual task, responses are selected in the semantic or in the premotor system.

sponses with response keys either labeled by a word or a color patch. For the task-response combinations of word-color and color-word, interference was obtained, as predicted by the translation model. Furthermore, also in accordance with the model, no interference was obtained in the word-word task. Contrary to the model predictions, Sugg and McDonald obtained interference in the color-color task with a negative stimulus-onset asynchrony (SOA), i.e., when the irrelevant word onset preceded the color onset (the interference disappeared at SOA = 0 ms). Thus there remains the possibility that the mapping of colors to keys involves some verbal mediation even if the keys are labeled by color caps. However, Durgin (2000) provided strong evidence for a translation account by changing the response format. When the response consisted in pointing to a matching patch of color, the Stroop effect disappeared, and instead a strong reverse Stroop effect was obtained.

Finally, there is also evidence that even early perceptual selection by visuo-spatial attention contributes to (the resolution of) Stroop interference. The critical results come from a color-block version of the task, in which word and ink color are spatially separated, by presenting a color bar above or below a word. Typically, Stroop interference is reduced in this version of the task, although it is still reliable if the spatial location of the target is uncertain, because its position is randomly drawn from trial to trial (Kahneman & Chajczyk, 1983; Hartley, 1993). However, under conditions of spatial certainty, interference is eliminated at SOA=0 ms (Glaser & Glaser, 1982), thus under certain[24] conditions the distractor can be filtered out by an system that implements early, visuo-spatial filtering.[25]

In summary, it is an open question at which stage of processing the conflict arises. However, it is becoming increasingly clear that Stroop interference is measurable at several levels (Sharma & McKenna, 1998; Roelofs, 2003; Peterson et al., 1999). As the response set and semantic gradient effects indicate, a number of different components contribute to interference, namely a lexical component (word vs. nonword), a semantic relatedness component (color-related word vs. color-unrelated word), a semantic relevance component (color word vs. color-related word), and a response set membership component. Thus interference occurs in at least the semantic, lexical, translation, and the response selection modules, each of which might recruit executive resources to overcome the conflict. The module at which the conflict is strongest may vary depending on the implementation details of the task.

Since on incongruent trials the conflict already arises on the input side of the cognitive stage, the Stroop color-word congruency manipulation, although ‘cognitive’, can still be regarded as a somewhat ‘early’ difficulty manipulation, at least if one compares it to the locus of the S-R compatibility manipulations that were discussed above.

Models of the Stroop task

Here I will more closely examine the processes that lead to Stroop interference. I will briefly discuss the Cohen et al. (1990) model as one prominent model of Stroop performance, and sketch an alternative class of dual-route models. The Episodic Accumulator Model, as applied to the Stroop task, inherits from both approaches. I will end this section by showing what patterns of interference effects are predicted by the episodic accumulator hypothesis in the Stroop task with and without an arbitrary S-R mapping.

The Cohen et al. (1990) model. Both empirical and theoretical work involves the concepts of controlled vs. automatic processing to explain Stroop performance. Of particular interest to the controlled/automatic distinction are the data obtained by MacLeod and Dunbar (1988), who trained participants to associate arbitrary shapes with a color name. After an initial learning phase where a color name had to be given in response to the presentation of a shape, the shapes were presented in a color that could be incongruent or congruent to the associated color. Early in training, incongruent colors caused large interference effects on shape naming, while incongruent shape-color associations caused no interference in a color-naming task. After a number of days of training, interference between the two tasks was symmetric, and after yet more training, the pattern reversed, so that incongruent shapes caused more interference on color-naming than incongruent colors did on shape naming. These results suggest that differential automaticity, or differential association strength, is one major source of Stroop interference.

Consequently, the feed-forward neural network model proposed by Cohen et al. (1990, see Figure 4) explains Stroop interference effects by assuming different association strengths of word and color pathways[26] with a reading response. Color and word pathways are distinct until they converge on a common response module representing the naming decision. For example, due to the stronger association of word input and reading output, presentation of an incongruent color word leads to a stronger activation in the word than in the color pathway, which would result in an erroneous word reading response. In order to give the correct response, external, ‘controlled attention’, executive task node inputs are needed, which differentially bias processing in the pathways to comply with task instructions. Processing in the word pathway is attenuated, while processing in

---

[24] pun not intended

[25] Posner and Petersen (1990) suggest that there are at least two distinct attention systems, a posterior and an anterior system. The posterior system involving parietal cortex, the pulvinar nucleus of the thalamus, and the superior colliculus can achieve visuo-spatial filtering, while the anterior system is concerned with attending to one of several possible streams of internal, cognitive processing.

[26] An extended version of the model also includes arbitrary shape input nodes.

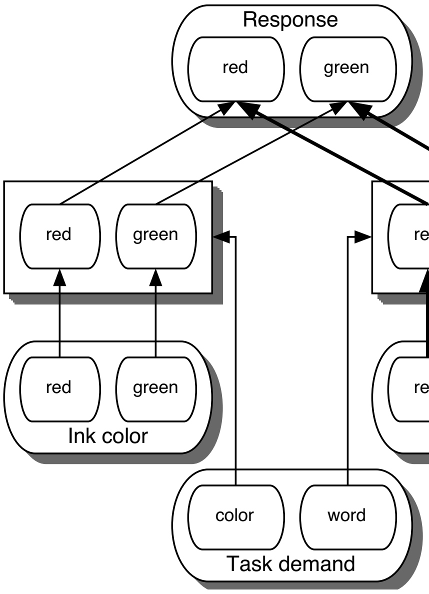


Figure 4. Simplified diagram of the feed-forward Stroop model developed by Cohen et al. (1990). Processing in separate color and word pathways converging on the same response set is modulated by top-down task demand input. Thicker lines indicate stronger associative connections. In addition to the between-layer facilitatory connections connecting associated elements, the model also includes between-layer inhibitory connections of equal strength, connecting red and green. The latter are not shown in the figure for reasons of clarity.

the color pathway is facilitated by inputs from controlled attention.

In the Cohen et al. (1990) model, differential association strengths in structurally identical pathways are the central explanatory mechanism, and the locus of interference is the response buffer, because there the two pathways meet for the first time. In recent extensions of the model (Botvinick et al., 2001), a conflict-monitoring module measures response conflict and feeds back an error signal into the cognitive control task input module. Thus a negative feedback loop is instantiated: the greater the conflict, the more control is called for, and greater biasing by attentional control in turn leads to reduced conflict. Interestingly, in the current implementation of this model, conflict-monitoring is not as dynamic as the other aspects of the model—the conflict signal is only fed back into the cognitive control module at the end of a trial, after a response has been given, while all other processing steps are repeated for many cycles during a single trial. Despite of these limitations, attentional control is no longer purely externally determined. Conflict signals the need for more control, thus more ‘resources’ are recruited for the differential modulation of processing in the pathways. Even the revised model, however, has no mechanism to explain interference at the semantic level, because it fails to distinguish between semantic and response selection levels. The model also has no mechanism to account for stimulus-ensemble specific, episodic repetition priming (see Mayr et al., 2003). Furthermore, it is difficult to see how the model can explain effects of switching the response modality.

Dual-route Models. The fast activation of an over-learned or highly compatible response by presentation of a stimulus is often explained by an alternative class of models, so-called dual-route models (e.g., Kornblum et al., 1990; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Neumann & Klotz, 1994; De Jong et al., 1994), which assume an architectural difference between compatible and arbitrary S-R associations. These models posit that presentation of a stimulus can activate the associated response via two parallel routes, a fast direct route, and a slower route that in the case of word reading passes through the lexical and semantic system. In the case of extremely over-learned or highly compatible S-R associations there exists a direct route from perceptual (e.g. graphemical) to response output (e.g., phonological) representations in addition to the semantically mediated route that has to be taken when less compatible or more novel associations are relevant. For example, the most successful model of reading aloud (Coltheart et al., 2001) assumes a direct, sublexical grapheme-to-phoneme conversion route in addition to the semantically mediated lexical route. In a Stroop task, the direct route leads to pre-activation of a response buffer (e.g., `/red/` phonology) in parallel with activation of the semantics (e.g., red color concept). In a reverse Stroop task (i.e., reading of the word while ignoring its color), there is no pre-activation of the wrong response, because color naming is not over-learned. In dual-route models, interference arises at both the semantic and the response level. In the case of an incongruent word in the color naming task, there is fast priming of the wrong phonological output buffer. The phonological output buffer for the color response can only be activated after the color has been semantically processed, by input from the semantic level. Because the wrong response is primed, more input from the semantic level is needed than in congruent or neutral conditions. Additionally, both the word and the color activate color concepts, so lateral inhibition and/or attentional control input originating from task nodes is needed to overcome conflict at the semantic level. Given a supra-threshold activation of the incongruent color concept by the distractor, activation cascades to the wrong response output, thus leading to conflict at the response buffer level. Again, conflicts can only be resolved by recruiting external resources. In the case of the word reading task, no reverse Stroop effect is observed, because response selection can be based entirely on the fast pre-activation of the word response. Conflict that still arises at the semantic stage arrives at the response buffer too late to significantly affect performance.

The WEAVER++ model developed by Roelofs as a model of spoken word production has recently been applied to the Stroop task (Roelofs, 2003). In this context, it can also be considered a dual-route (or multiple-route) model. When modeling the Sugg and McDonald (1994) data with WEAVER++, Roelofs made the assumption that responses are generated using the shortest route possible from stimulus to response while respecting the response type. According to this model, Stroop interference always arises at response selection, but the location of the system where responses are selected depends on the task at hand. In the vocal task, responses are selected in the lexical system. In the manual task, response nodes are connected to concept nodes for color responses, to lemma nodes for color-driven word responses, and to form nodes for word-driven word responses. Thus although manual responses are involved in all three conditions, how lexical entries mediate the response critically differs between conditions–the level at which the selection is made is critical for performance in the manual version of the Stroop task.

Arbitrary rules in the Stroop task

One prominent specific-process theory of aging posits that old adults have difficulty inhibiting irrelevant information in working memory (Hasher & Zacks, 1988; Hasher et al., 1991; Kane et al., 1994). This theory has been criticized mainly because of the diffuse concept of inhibition (see the discussion in the Journals of Gerontology, 52B(6): Burke, 1997; McDowd, 1997; Zacks & Hasher, 1997). In fact, depending on how inhibition is measured, over-proportional age-difference have been observed or not. As has been argued above, resolution of Stroop interference requires active inhibition of the incorrect response. Thus according to the inhibitory-deficit hypothesis, large age effects can be expected in a Stroop task. However, as a recent meta-analysis shows, the slowing factor in the congruent and incongruent conditions seems to be equivalent, indicating that active inhibition in Stroop-like tasks is not specifically slowed (Verhaeghen & De Meersman, 1998).

Here I argue that the degree to which age-differences in Stroop interference are found critically depends on the degree of arbitrariness of the task rules in a Stroop task. In the case of compatible stimulus-response mappings, I predict age effects to be relatively small, while larger age effects are expected when nonstandard mappings are used. This is because interference resolution takes place in working memory (or the episodic buffer stage) in the latter case, while it requires less working memory capacity in the former.

What happens to the sources of interference in the Stroop task if an arbitrary S-R mapping is used? As has been argued above, interference in a Stroop task arises at several levels. In the case of arbitrary S-R mappings, interference by direct activation of the compatible or over-learned associated response is impossible or at least weak—only interference at earlier, e.g. lexical and semantic levels, is relevant. Thus interference arises at the input side of the arbitrary rules, which map semantic color concepts to either phonologically or spatially coded response-related concepts. Because there is competition at the input side of the rules, I consider the manipulation of Stroop interference an ‘early’ difficulty manipulation relative to the manipulation of mapping.

A standard result is that the magnitude of the Stroop effect is smaller with manual than with verbal naming responses (White, 1969; Redding & Gerjets, 1977; Logan et al., 1984; Sharma & McKenna, 1998; Keele, 1972; for a summary see MacLeod, 1991). This is often explained by the fact that the representation of the key is not pre-activated by the word distractor, i.e. there is no direct route from stimulus to response. The executive, working memory demands of an arbitrary mapping (e.g. color → key) have been neglected in previous discussions of this pattern. One reason for this neglect might be the fact that a resource or capacity conception of controlled attention would actually predict larger interference in the manual response condition, because the maintenance of the arbitrary mapping requires additional resources from the same pool that inhibition of the distractor draws from. This apparent paradox can be resolved by recognizing that there is a direct route from word input to naming output, and that this direct route from distractor to response is absent when using manual responses.

Thus there might be two partially opponent processes contributing to Stroop interference. On the one hand, pre-activation of the wrong response is the higher the greater the degree of dimensional overlap between distractor and response (and the smaller the degree of dimensional overlap between target and response—see Kornblum et al., 1990, for a treatment of the concept of dimensional overlap). If distractor and response codes overlap, then there is a sub-semantic pre-activation of the response by the distractor along the direct route. Executive resources are needed for the active inhibition of the pre-activated wrong response. The amount of resources required will be related to the degree to which distractor and response codes are more similar to each other than are target and response codes. Evidence for this process comes from the comparison of the Stroop effect and the reverse Stroop effect, which is typically small or absent. On the other hand, the direct route is absent (or at least weak) if manual responding is used in a Stroop task, because neither reading nor color perception are strongly associated with pressing a key. Thus executive resources are not needed for active inhibition of a pre-activated response. Rather, they are needed (a) for inhibition of the semantic concept activated by the distractor, and (b) for maintenance of the arbitrary rules.

In a standard Stroop task using vocal responding, there are no arbitrary S-R mappings—rather, the mapping of color concept to naming response is compatible. However, the S-R associations between target and response are weaker than those between distractor and response[27]. Thus arbitrary rules in the standard Stroop color naming task are not very eminent (although one arbitrary task rule might consist in following the instructions not to react to the word, but to its color). In a Stroop task with manual responses, an additional set of

---

[27] In other words, the word → naming relationship is more compatible than the color → naming relationship.

arbitrary rules is introduced by the mapping of color semantics to keys. Therefore, the overall degree of arbitrariness is greater in the manual Stroop task. Hence, relatively more executive resources than in a standard Stroop task will be needed for the maintenance of the arbitrary mapping and for protection from interference in working memory, while relatively fewer resources are required for protection from interference at the response selection level.

To summarize, in the case of arbitrary S-R mappings, Stroop interference can not be caused by pre-activation of a response via a direct, automatic route. Instead, a semantic origin of interference is likely. As long as the same stimuli are used, Stroop tasks with compatible and with arbitrary S-R mappings do not differ in the degree of interference at the semantic level, simply because the same concepts are activated. With a compatible mapping and a relatively large degree of overlap of distractor and response code, additional interference arises at the response selection level due to the direct S-R route.[28] While direct interference at the response selection level cannot occur in arbitrarily mapped tasks, a new locus of interference is in the module that is responsible for maintenance of the stimulus-response bindings (or for 'translation' of the color semantics into a spatial code), i.e. the 'episodic' working memory component of the task.

Within the dual route model of Stroop interference sketched above, the standard result of reduced Stroop interference in the case of manual responding is easily explained, if it is assumed that young adults might not have much of a problem in maintaining the bindings (i.e. they have reliable episodic accumulators), even in the case of co-activation of two rules arising through activation of two concurring concepts at the semantic level. Thus, that part of interference that is caused by the episodic buffer-type short-term memory does not matter much for them, in comparison to the larger part of interference caused by the fast priming of the wrong response along the direct route when responding vocally. Since the latter does not play a role when responding manually, Stroop interference is reduced in comparison to vocal responses. It is not completely absent, because semantic interference is not totally negligible.

The episodic accumulator hypothesis posits that age-related deficits in the maintenance of arbitrary task rules are a likely specific cause of age-related slowing. If this is true, then old adults should show relatively high levels of interference in a Stroop task using arbitrary S-R mappings. Because they suffer from interference in working memory, old adults should profit less than young adults from the absence of overlap between distractor and response. In principle, it is conceivable that in old age the costs due to episodic buffer interference are larger than the benefits due to reduced response competition as a consequence of the omission of the direct route.

Sketch of an Episodic Accumulator processing model for the Stroop task

To more formally investigate the predictions by the Episodic Accumulator Model in a Stroop task, I developed an interactive activation model that combines the idea of dual-route models for the Stroop task with the Episodic Accumulator Model. The model is depicted in box and arrow form in Figure 5 and is implemented as an interactive activation, feed-forward model.[29]

There are two routes from stimulus input to response output, one semantic route that is activated by all stimulus inputs, and an 'automatic' direct route from the stimulus to the response layer that is only available for word stimuli. The architecture of the semantic route of the model is similar to the Cohen et al. (1990) model, i.e., there are word and color input nodes as well as task input nodes, which implement attentional control. The major difference is that the response nodes in the Cohen et al. model are interpreted as semantic classification nodes in the current model. Different from the Cohen et al. model, automaticity of word reading is not modeled by (widely) different activation strengths in the word and color pathways, but by the addition of the direct route from grapheme encoding to the response layer, which was inspired by the DRC model of reading aloud (Coltheart et al., 2001). Critically, all aspects of the model that have been discussed so far assume age equivalence.

The Episodic Accumulator Model is attached to the output of the semantic nodes. The model is identical to the version described above (pp. 15 ff.) and is therefore another example of a dual-route model (embedded in the larger dual-route model). Within the Episodic Accumulator Model, there are direct connections from input to response in addition to the indirect pathway via episodic accumulators. In the indirect pathway, input is fed into episodic accumulators, which in turn project to response nodes. In both cases, input is constituted by output activation in the semantic nodes of the Stroop model. Within the Episodic Accumulator Model, compatibility of input and response representations determines the relative weights of the paths connecting inputs to response and to episodic accumulators. If input and response are compatible (e.g. say "blue" in response to the concept blue), then the weights in the direct path are relatively

---

[28] For example, graphemical encoding of the distractor word and phonological encoding of the verbal response overlap, so that presentation of a word leads to a fast reading response. Graphemical encoding and the spatial code used for manual responding do not overlap.

[29] Since this is only a 'toy' model, implemented to show that the qualitative pattern of results can be obtained as stated in the text, no attempt was made at fitting parameters using an automated procedure like gradient descent or a genetic algorithm–rather, parameters were chosen by hand and intuition. Furthermore, different from the Cohen et al. (1990) model, the current model does not include a learning mechanism, such as error backpropagation, because I did not intended to simulate learning in the task.

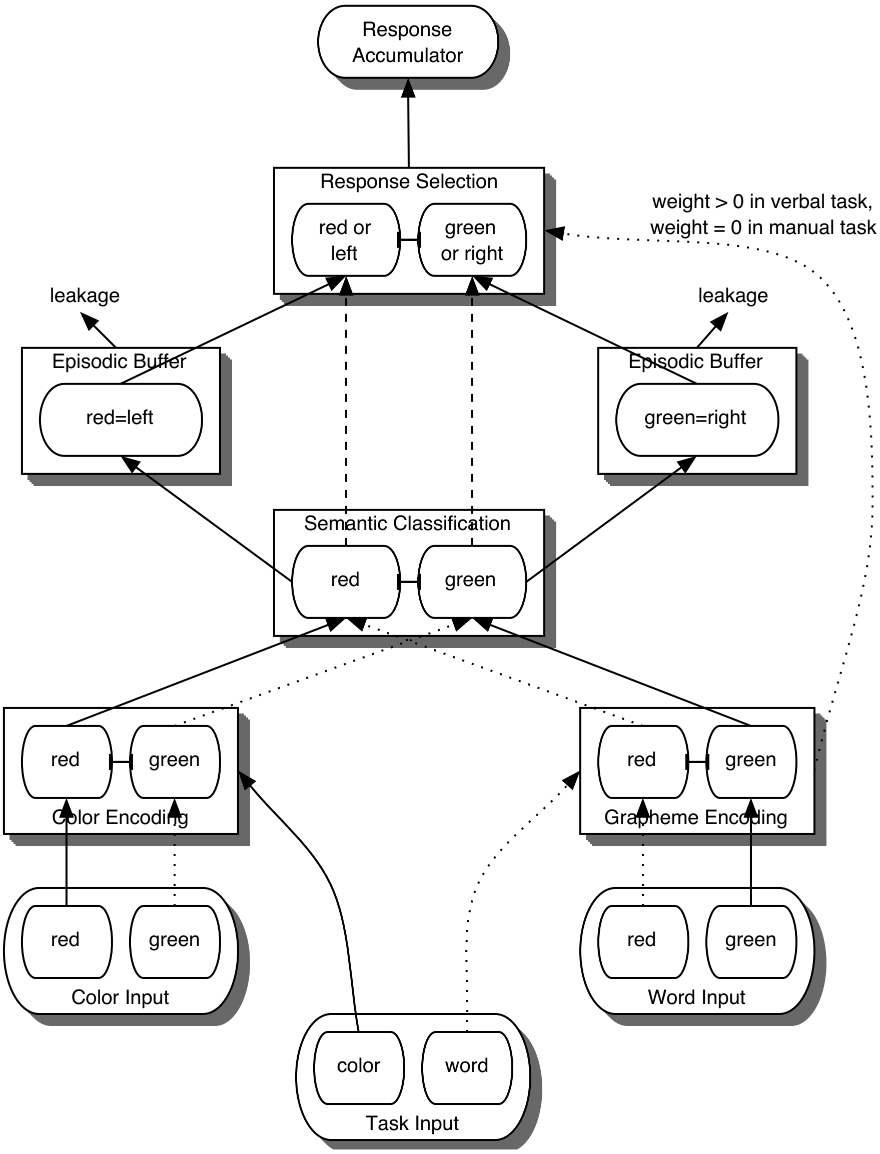


Figure 5. Combination of a model of the Stroop task and the Episodic Accumulator Model. The Stroop model (lower part of the figure) was inspired by the well-known model by Cohen and colleagues (1990). The Episodic Accumulator Model (upper part of the figure) has been described above. A direct route is added from graphemical encoding of the word to the response selection layer in the case of vocal responding.

stronger than in the episodic path, while in the case of an arbitrary mapping (e.g. press left key in response to the concept blue), weights in the episodic path are stronger than in the direct path. Thus input-response compatibility is a relative concept, determining the relative reliance on the direct and the episodic pathway.[30]

The following paragraphs will give a brief sketch on how processing in the model is implemented. There are separate sets of input nodes for color and word information. Each input set consists of two nodes, one coding for red and the other for green. For example the word red printed in green color has a word input pattern of [red=1, green=0] and a color input pattern of [red=0, green=1]. Additionally, there is a set of ‘controlled attention’ input nodes that code for the task at hand, i.e. word or color. In the present simulations, only the color task input node was active. Input to units in all other layers is weighted by connection strengths and summed up to give the net input into a node per cycle. The current activation of a unit is determined by first calculating the current net

---

[30] In the arbitrary case, weights in the direct path could be set to zero, however they are kept at a small value above zero to account for the fact that some degree of automaticity develops during the course of an experiment. *Development* of automaticity is not modeled, because it is not the primary focus here. The current set of parameters may thus correspond to a trial in the middle of a single-session experiment, where some automaticity has developed, but the association between concept and response is still relatively weak.

input at time t for unit j

$$net_j(t) = \sum a_i(t)\, w_{ij} \qquad (11)$$

where $a_i(t)$ is the activation of each unit the current unit receives input from, and w_{ij} is the strength of the connection of unit i and unit j. (The architectural connectivity in the model is shown in Figure 5 and has been discussed above.) Next, a weighted sum of the current net input and the running average of the net input in previous cycles is calculated, as in the Cohen et al. (1990) model (p. 337):

$$\overline{net}_j(t) = \tau net_j(t) + (1-\tau)\overline{net}_j(t-1) \qquad (12)$$

where $\overline{net}_j(t)$ is the time average of the history of net input to unit j at time t, and τ is a rate constant. Finally, the logistic function is applied to $\overline{net}_j(t)$ to determine the current activation of unit a_j:

$$a_j(t) = \frac{1}{1+exp(-\overline{net}_j(t))} \qquad (13)$$

At the output end of the model, response accumulators collect evidence for each response alternative. These sum up the difference in activation of the associated response node and the most active alternative. Because currently only two response accumulators are used, this is the difference between the associated response node and the response node associated with the other accumulator. The cycle at which a fixed response threshold is first reached in one of the accumulators is noted as the response time of the model (and the identity of the accumulator determines the response identity). Although it would have been easy, I decided not to add random noise to the response accumulators (or at any other stage of the model) because I only wanted to simulate a qualitative pattern of mean response times.

How does information flow in the model? Consider as an example an incongruent color-naming trial, (a) with a vocal, and (b) manual response. When responding vocally, after word and color input are encoded word information directly reaches the response selection stage, thereby priming the wrong response. Along a parallel route, word and color information reach the semantic classification stage, where both concepts are activated. However, due to top-down task input, activation of color input is enhanced relative to word input. When the activated concepts are colors, and the response is vocal, there is a relatively strong direct connection from semantic classification to response selection. However, when responding is manual, this direct concept-response connection is weak. Parallel to the direct semantic concept-response route, semantic stage information is fed into the episodic buffers, which implement the arbitrary mapping. This path is particularly important if the concept-response association is arbitrary, e.g., with manual responding to colors. Old adults' episodic buffers are assumed to be less reliable than young adults', thus they leak more information. In effect, old adults take longer to overcome the semantic conflict in the manual response condition. To summarize, in the vocal-response task interference arises through direct activation of the verbal response by graphemic encoding, and is resolved mainly via the direct semantic-response route. In the manual-response task, interference mainly arises at the semantic level and is resolved along the episodic buffer route.

What predictions does this model make, (a) for a Stroop task with compatible responses, and (b) for a Stroop task with arbitrary responses? How does reduced reliability of episodic accumulators in old age affect the Stroop effect? To investigate this, model predictions for a total of eight conditions were generated for the color naming task. The factors were Age (young vs. old), Mapping (compatible vs. arbitrary), and Stroop condition (color-word congruency: congruent vs. incongruent). Results are presented in Figures 6 and 7. It turns out that a triple interaction of Age, Mapping, and Stroop condition is predicted. In all cases, if compared to a group of simulated 'young' subjects with reliable episodic accumulators, the 'old' group (with unreliable episodic accumulators) showed relatively large Stroop effects in arbitrary mapping conditions. The size of this effect depends on the relative strength of the direct route. Recall that one motivation to add a direct route was the standard result for typical subjects, i.e. young adults, that interference effects are smaller with manual than with verbal responses. In the simulations, if a sufficiently weak direct route weight was chosen, the size of the Stroop effect for the old group was actually even larger with an arbitrary than with a compatible mapping (Figure 7 (left panel)). On the other hand, with a sufficiently strong direct route weight, even the old group produced larger Stroop effects with a compatible than with an arbitrary mapping (Figure 7 (middle,right panel)). Thus, if the model architecture is assumed to be correct, whether or not the Stroop effect in the arbitrary mapping is larger or smaller than in the compatible mapping condition is not diagnostic with respect to the question of the reliability of episodic accumulators. Rather, the critical between-group contrast is the relative change in Stroop effect when comparing the compatible and arbitrary mapping conditions.

Experiments

Overview

In this section, I will give a short overview of the experiments. Experiment 1 combined manipulations of perceptual difficulty and stimulus-response compatibility (SRC) in a classic SRC paradigm using manual responses delivered on a touch-screen. Experiments 2-6 made use of the Stroop task as cognitive difficulty manipulation and used different variants of the task to vary the degree of arbitrariness of task rules. In most experiments, episodic demands were manipulated by varying the complexity of the set of rules mapping stimuli to responses. In particular, all experiments had a high-demand condition using a set of arbitrary mapping rules.

In stage theories, the SRC manipulation is considered to affect the response selection stage. In the

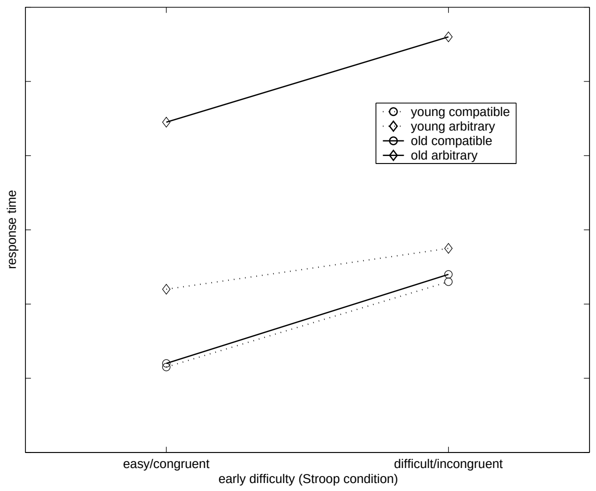


Figure 6. Response times predictions of the dual-route Episodic Accumulator Model in the case of cascaded processing and input from a model of semantic activation in the Stroop task.

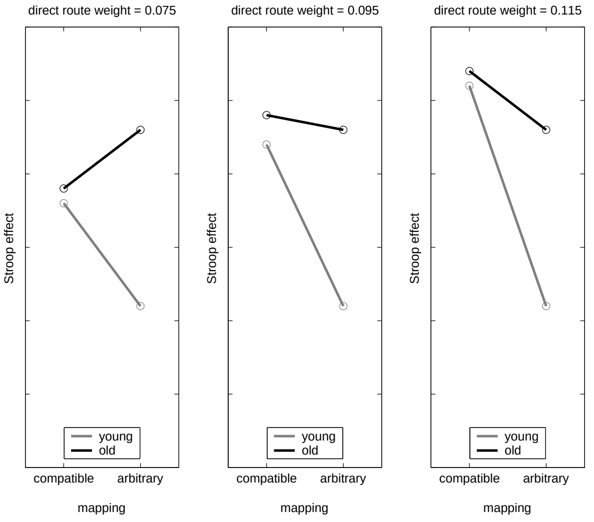


Figure 7. Stroop effect predictions of the dual-route Episodic Accumulator Model in the case of cascaded processing and input from a model of semantic activation in the Stroop task. From left to right, the relative weight of the direct stimulus-response route increases.

first experiment, stimulus discriminability was manipulated as an additional factor which—in contrast to the SRC manipulation—affects early stages of information processing, namely stimulus encoding and recognition. Due to the relative independence of stimulus classification and response selection stages, it is likely that effects of difficulty manipulations affecting the former will be independent of effects of manipulation affecting the latter. If there were no slowing of early information processing, age and perceptual difficulty should produce additive effects. However, a more realistic assumption is that there are measurable age effects of the early difficulty manipulation, but that these are relatively small. For later stages of information processing, the episodic deficit hypothesis predicts an interaction of age and SRC, more specifically, large age effects in the case of arbitrary mappings. The question remains whether age effects in early stages are amplified by age effects in a later, 'episodic' stage. The Episodic Accumulator Model predicts such an over-additive triple interaction of age, early difficulty, and SRC only when episodic buffers start receiving information before 'early' classification is finished. Thus if a triple interaction were obtained in Experiment 1, then the relative independence of stages—as posited by the additive factors program—would have to be questioned in the context of aging research. On the other hand, if the interaction were not obtained despite of age differences due to the 'episodic' manipulation, then it can be concluded that the hypothetical episodic buffers are located at a relatively late stage of information processing, after perceptual classification.

Experiments 2-5 added an SRC manipulation to the Stroop task and differed in the way stimulus and response sets as well as the arbitrary mapping rules were implemented. The Stroop manipulation was chosen as a manipulation of 'cognitive difficulty' at the input side to the episodic buffers. Because conflict in the Stroop task is promoted up to the response selection stage, a critical three-way interaction of Age, Stroop condition, and Mapping (episodic demands) was predicted by the hypothesis of reduced reliability of episodic accumulators. Experiment 2 used color-word stimuli with manual as well as vocal responses, and additionally addressed the question whether integration of stimulus and distractor on a single perceptual object critically mediates age differences. Experiments 3 and 4 used color-word stimuli and vocal responses, comparing compatibly-mapped with different sets of arbitrarily-mapped concepts. In Experiment 3 a set of male first names was mapped to colors, and in Experiment 4 the set of cardinal numbers one to four was used. Experiment 4 also added a manual response condition that replicated the manual condition of Experiment 2. To address the question whether 'episodic' age effects are specific to or at least more pronounced in the manual response modality, Experiment 5 used a spatial Stroop paradigm that allowed for the orthogonal manipulation of episodic demands and response modality. Finally, Experiment 6 investigated whether differential slowing in the spatial and verbal domain could have led young and old adults to use different strategies for the coding and maintenance of the arbitrary mapping. Additionally, Experiment 6 inspected the reverse Stroop effect in the case of manual responses and possible age differences therein.

General Methods

To avoid repetition, before the actual experiments are presented I will shortly describe the methodological aspects common to all of the experiments reported below.

Participants. Participants were drawn from the University of Potsdam cognitive psychology department subject pool. Participants were paid for their participation, except for a subgroup of young adults consisting of undergraduate psychology students, who had the option to receive course credit instead of financial reimbursement. Young adults received DM 10.00 (about $ 5.00) per hour, and old adults received DM 15.00 (about $ 7.50) per hour.

All participants were healthy according to self-report and had normal or corrected-to-normal vision, as assessed by a Snellen acuity chart. At the beginning of each experiment, participants filled in a demographic questionnaire, and two tests were administered, both of which constitute routinely applied measures in cognitive aging research, and are derived from the Wechsler Adult Intelligence Scale - Revised (WAIS-R, Wechsler, 1981, cf. Lindenberger et al., 1993). The "Zahlensymboltest" (Tewes, 1991/1994) is the German version of the digit-symbol substitution (DSS) subscale of the WAIS-R, which is considered a measure of "psychomotor speed". Ten symbol-digit (key-value) pairs are given on the top of a sheet of paper, and the actual test requires writing down the matching digit for a number of symbol items under speed pressure. The "Mehrfachwahl-Wortschatz-Intelligenztest" (MWT, Lehrl, Daun, & Schmidt, 1971; Lehrl, 1977/1995) is a German multiple-choice version of the WAIS vocabulary subscale. It measures vocabulary knowledge as a prototypical form of crystallized intelligence. We used the version MWT-A (Lehrl et al., 1971), which is a parallel form of the MWT-B (Lehrl, 1977/1995). Each item of the MWT requires underlining the one proper German word from five simultaneously presented character strings (e.g. *Amarika - Akarina - Amakira - Amaki - Amerika*). Items are ordered by increasing difficulty, and there is no speed pressure. The digit symbol substitution test is finished after 90 seconds, and the MWT-A requires about five minutes on average.

Roughly, in differential psychology terms, the digit symbol substitution test measures fluid intelligence (sensu Cattell), while the MWT-A measures crystallized intelligence. From a cognitive psychology perspective, it could also be argued that MWT-A is mainly a semantic long-term memory test, while digit symbol substitution taxes attention and short-term/working memory. The attentive reader of the introductory chapter will thus predict that age differences should be relatively small in the MWT-A, and substantial in the DSS. This is what is typically observed, in fact, older adults' performance in the MWT-A is sometimes even better than young adults'.

Apparatus. The experiments were controlled by a Power Macintosh 7500 computer running Mac OS 8.6, which was programmed in C and Pascal using in-house developed libraries and routines from the VideoToolbox library (Pelli, 1997).

A touch-screen was used in Experiment 1, and will be described in greater detail in the according section. The following is valid for all other experiments. All stimuli were displayed on a 17" Apple monitor with a resolution of 832 x 634 pixels at a vertical refresh rate of 75 Hz. Stimulus display was synchronized to the vertical blanking interrupt. Viewing distance was about 70 cm. Vocal responses were collected with a Sony condensator microphone connected to the CMU button box, which has a built-in voice key with a temporal resolution of 1 millisecond. At the beginning of each session that involved vocal responses, the sensitivity of the voice-key was calibrated to match individual participants' loudness. Calibration was performed by asking participants to read aloud words presented on the screen, while the experimenter manually adjusted the sensitivity. Vocal responses were also recorded on DAT tape using a second microphone. During the experiment, the experimenter who was sitting behind the participant monitored participants' responses in the vocal condition and noted errors on a prepared sheet of paper that listed the full sequence of stimuli and normatively correct responses for all vocal trials. After the experiment, a second experimenter reviewed the DAT recording and coded errors on a second, identical sheet of paper. Three categories were used for error coding: correct response, wrong response, and error due to technical problems. The latter could include voice key failures due to muttering, external noise, insensitivity, etc. Manual responses in experiments with only two response alternatives were collected using the built-in keys of the CMU button box, while manual responses in experiments with four response alternatives were collected on the Apple keyboard connected via ADB. While the button box provides a higher temporal resolution (1 ms vs. 12-16 ms), it has only three response keys and could thus not be used in the latter type of experiments.[31]

Data analysis. For all reaction time ANOVAs, data from error trials and trials following an error were discarded. Two outlier removal criteria were then sequentially applied. First, an absolute maximum threshold, which varied from experiment to experiment, was applied to remove extreme values which might otherwise bias the linear aggregate statistics. Second, RTs that deviated by more than 3 SD units from their design cell mean per participant were discarded. Raw data were then aggregated across trials per design cell and participant, using the arithmetic mean as a summary statistic. Aggregated RTs were submitted to repeated measures ANOVA.

Further analyses were performed to take proportional slowing into account. General slowing leads to the

---

[31] The relatively coarse granularity of the timer in experiments with four manual response alternatives does not appear to pose a major problem. As long as a reasonable number of measurements per mean is taken, the mean is unbiased, and the granularity of the timer is only reflected in the variance (Ulrich & Giray, 1989).

expectation that the average response latency for old adults, minus the time taken for peripheral processing, is a constant proportion of the average response latency for young adults, if peripheral factors are subtracted (see equation (7)). Thus, if mean reaction times are analyzed, spurious over-additive interactions of age group and treatment can be expected based on general slowing alone. Thus additional analysis were performed with different measures in an attempt to control for baseline speed differences. These include (a) an analysis of logarithmic reaction times, and (b) an analysis of a proportional interference score in Experiments 2-6, namely, the difference of latencies in the incongruent and congruent conditions, normalized by the baseline latencies. Unless otherwise reported, proportional analyses did not qualitatively alter the results of the raw RT significance tests.[32]

Experiment 1

Age effects in an S-R compatibility paradigm

Are 'perceptual' age differences amplified by an unreliable representation of stimulus-response rules? According to the episodic accumulator hypothesis, over-additive interactions of age early difficulty are expected when it is difficult to maintain the representation of the S-R mapping, relative to a situation in which episodic accumulators can be by-passed. However, if the episodic accumulators are located at a late stage, and difficulty manipulations affect an early, perceptual evaluation stage, then early and late difficulty manipulations affect different discrete stages. Because information flow between discrete stages is by definition non-cascaded, additive effects of early and late difficulty manipulations are expected.

This hypothesis was tested in a reaction time paradigm, where participants had to touch fields on a touch-sensitive screen in reaction to spatially oriented stimuli. In addition to age, there were two factors. First, perceptual difficulty was varied by changing the relative contrast of a target square and five distractor squares, thereby varying target discriminability, making the target either easily detectible or hard to detect. The contrast manipulation was implemented by changing the amount of randomly distributed white pixels per square. Second, difficulty of the S-R mapping was varied in three steps. The mapping was either compatible, i.e. the deviating (target) square had to be touched, incompatible, i.e. the square that was horizontally opponent to the target square had to be touched, or arbitrary, i.e. the mapping between stimulus and response locations did not follow a simple rule. Because the rule in the incompatible condition is rather easy to code, only the arbitrary mapping condition constitutes a situation with high demands on the system representing task rules. However, inclusion of the inverse rule allows for a test an alternative account of deviations from general slowing, namely the 'inhibitory deficit' idea (e.g., Lustig et al., 2001) claiming that old adults are particularly affected in situations that require the inhibition of dominant reactions (e.g., touching the compatible square).

Presumably the direct, compatible S-R mapping puts no load on working memory, in other words episodic task components are very low. Working memory demands are only little higher in the inverse condition, since the mapping rule can be generated from the direct mapping by the simple transformation rule "touch the (horizontally) opposite square". In the arbitrary condition, however, episodic task demands are considerably higher, since the mapping rule cannot be generated from the natural mapping by such a simple transformation—instead, activation of the correct response has to proceed by a process akin to table lookup. The arbitrary set was designed to include element-level rules from the inverse set for a number of elements. This constraint was introduced to allow for comparisons of single elements between the inverse and the arbitrary set, which had exactly the same mapping rule (e.g. upper right→upper left). Differences in reaction times to those elements between mapping sets thus measure pure set effects and can be interpreted as effects of episodic task difficulty.

Method

Participants. Eighteen younger (age $M = 19.6$ years, *range* $= 17 - 24$) and 18 older adults ($M = 69.8$, *range* $= 65 - 77$) participated in this experiment. Both age groups were approximately equal with respect to years of formal education (young, $M = 12.8$, $SD = 2.8$; old, $M = 12.2$, $SD = 1.8$), $t(34) < 1$. Age groups showed the usual pattern regarding fluid intelligence/mental speed and crystallized intelligence/semantic knowledge, with young adults scoring higher on the digit-symbol substitution test (young, $M = 60.7, SD = 6.5$; old, $M = 50.7, SD = 10.0$), $t(34) = 3.58$, $p = .001$, and old adults scoring higher on the MWT-A vocabulary test (young, $M = 30.3$, $SD = .64$;

---

[32] In principle, the recommendations given by Faust, Balota, Spieler, and Ferraro (1999) could have been followed. These authors compared proportional or logarithmic transformations, z-score transformations, and regression transformations. The latter involve simple linear regression of the overall means for each condition on each individual's latencies. The resulting regression predictions can then be analyzed. Faust et al. recommend the use of regression predictions when applicable (i.e., if predictions are based on "a wide enough range of individuals and conditions", Faust et al., 1999, p. 793), and the use of z-score transformations otherwise. The use of proportional scores is not recommended, because it is biased in the case of a Brinley plot that does not pass through the origin (but the intercept of a young-old Brinley plot is typically negative). One problem with the recommended transformed value analyses is that they are rather difficult to interpret and to compare with untransformed reaction time analyses. For example, it is clear that the age main effect in z-score and regression transformation must be absent. However, Stroop experiments allow the calculation of an easily interpretable proportional measure, namely the proportional interference score calculated by (incongruent-congruent)/baseline latencies. Because of its advantage with respect to interpretability, this measure was used in Experiments 2-6.

old, $M = 32.7$, $SD = .39$), $t(34) = 3.11$, $p = .004$. All subjects were healthy according to self-rating and had normal or corrected to normal vision.

Stimuli and responses. The stimulus display consisted of two columns of squares with a side length of 80 pixels (3° VA) each, which were defined by white borders one pixel wide on a black background. Each column consisted of three evenly spaced squares, arranged inside a centrally presented virtual rectangle of 170 by 260 pixels (6.4 by 9.8° VA when viewed from a distance of 60 cm). A square of the same size as the stimulus-response squares was located centrally .5 degrees below the two columns and used as a 'finger fixation device'.

Perceptual difficulty was manipulated by changing the difference in brightness between squares. All of the squares were filled with random dot patterns. While all but one of the squares had the same density of random dots, with 10% of pixels inside the square lit, the density was changed for the remaining square, which I will call the imperative stimulus (IS). In the perceptually easy condition 70% of IS pixels were lit, while in the perceptually difficult condition only 15% were lit. The IS was the response-relevant part of the stimulus display. The responses were given by touching regions of a touch-sensitive screen with the pointing finger of the dominant hand. Touch-sensitive regions were defined by the stimulus squares. Subjects were instructed to touch a square on the screen which corresponded to the deviating one. The correspondence relation was defined by a set of mapping rules. Each rule set provided element-level mapping rules for locations of the deviating stimulus onto locations of the desired response. Three different sets of mapping rules were used: compatible, inverse and arbitrary mappings. In the compatible mapping condition subjects simply had to touch the deviating square. In the inverse mapping condition, subjects had to touch the square in the opposite column, e.g. if the deviating square was in the upper right location, the desired response was a touch of the upper left square. Thus for the indirect set the mapping rules were [(vertical position) left→(vertical position) right; (vertical position) right→(vertical position) left], where vertical position is one of upper, middle, lower. The arbitrary mapping condition was had a more complex set of rules to increase working memory demands (or episodic difficulty) of the task. For half of the subjects the arbitrary set was [upper left→upper right; middle left→lower right; lower left→middle right; upper right→lower left; middle right→middle left; lower right→upper left]. For the other half of the subjects, the arbitrary mapping was mirrored between columns.

Design and Procedure. The design was a 2×2×3 mixed factorial, with the between-subjects factor Age (young, old) and the within-subjects factors of Perceptual difficulty (easy, difficult) and Mapping (compatible, inverse, arbitrary).Mappings were blocked, with order balanced between subjects and matched between age groups. The relevant set of mapping rules was presented at the beginning of a block pictographically (see Figure 8 for examples depicting inverse and arbitrary mapping rules). Participants were encouraged to study the picture until they felt they could remember the rule set well Each block consisted of 64 training trials and 288 test trials amounting to a total of 192 training trials and 864 test trials. Training trials were included to acquire the mapping rules. They did not otherwise differ from test trials, but they were not included in the analysis. Perceptual difficulty and target location were randomly drawn (without replacement) on each trial within a block. After every 64th trial, subjects could choose to pause.

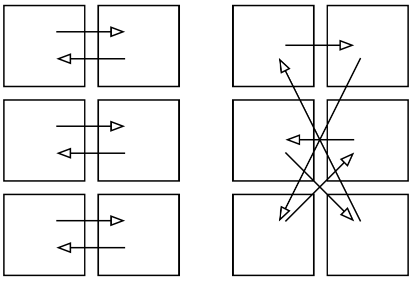


Figure 8. Inverse (left) and arbitrary (right) mapping rules in Experiment 1.

A trial started when participants had moved their finger to the finger fixation square. First, the six stimulus squares and the fixation square were presented as empty outlines. After 1000 ms, to give a warning signal for the upcoming stimulus, the finger fixation square flickered for 156 ms[33]. A further 200 ms later, the random dot patterns were presented inside the stimulus squares and remained visible until a correct response was given by touching the square that corresponded to the IS according to the element-level mapping rule. Thus, implicit feedback was given, because the next trial was only started after the correct square had been touched, in other words, errors had to be corrected to proceed with the experiment.

Results and Discussion

Reaction times. After removal of outliers (1.3%), aggregated cell means (see Figure 9) were analyzed using a repeated measures analysis of variance with age as between-subjects factor and the two within-subject factors of Mapping and Perceptual difficulty. All main effects were significant: old adults were slower than young adults (Ms= 902 vs. 690 ms), $F(1,34) = 33.31$, $p < .001$, $MS_e = 72566.63$, the Perceptual difficulty manipulation was successful (difficulty effect 90 ms), $F(1,34) = 179.43$, $p < .001$, $MS_e = 2433.35$, and a compatibility effect was obtained, $F(2,68) = 424.63$, $p < .001$, $MS_e = 26528.06$.

Single comparisons revealed that responding with the compatible mapping ($M = 534$ ms) was faster than

---

[33] This involved three repetitions of deletion and presentation of the square, with a presentation duration of two frames each.

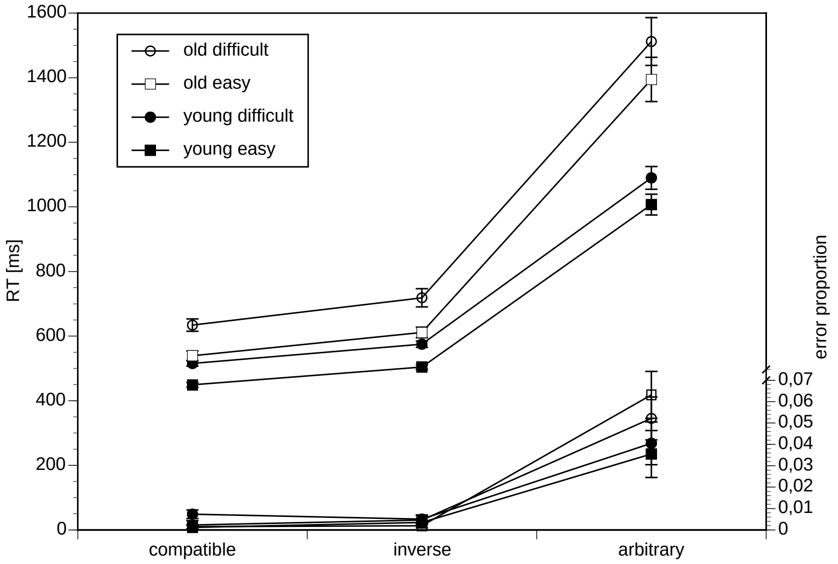


Figure 9. Experiment 1: Reaction time (upper set or curves, left value axis) and errors (lower set of curves, right value axis), split up by age, perceptual difficulty, and stimulus-response compatibility (mapping). Error bars indicate standard errors of the mean.

with the inverse mapping ($M = 602$ ms), $F(1,34) = 86.36$, $p < .001$, $MS_e = 3830.40$, which in turn was a lot faster than responding with the arbitrary mapping ($M = 1251$ ms), $F(1,34) = 407.21$, $p < .001$, $MS_e = 74402.25$).

Mapping and Perceptual difficulty did not interact in raw reaction times, $F(2,68) = 2.55$, $p = .086$, $MS_e = 684.59$. There was a small increase in the Perceptual difficulty effect from compatible via inverse to arbitrary mapping, however, the interaction was not significant in either of the repeated contrasts used for pairwise comparisons. Note that an under-additive interaction appears in log reaction times (see below), which suggests that this increase is under-proportional.

Age interacted with Perceptual difficulty, $F(1,34) = 6.14$, $p = .018$, $MS_e = 2433.35$: the Perceptual difficulty effect was larger for old adults than for young adults (106 vs. 73 ms), however, the lack of significance in the analysis of log reaction times (see below) indicates that this effect can be accounted for by proportional slowing.

A substantial interaction of Mapping and Age was obtained, $F(2,68) = 19.06$, $p < .001$, $MS_e = 26528.06$. Single comparisons reveal that the age effect was larger with the arbitrary than with the inverse rule, $F(1,34) = 18.88$, $p < .001$, $MS_e = 74402.25$, while for the comparison between the compatible and the inverse mapping, the effects of Age and Mapping were additive, $F(1,34) = 2.04$, $p = .163$, $MS_e = 3830.40$. Because responding in the inverse mapping requires inhibition of the dominant response, namely responding with the compatible mapping, the finding of an absence of age differences in the compatible-inverse comparison is difficult to reconcile with (extreme versions of) an inhibitory deficit theory.

There was no three-way interaction of Age, Mapping, and Perceptual difficulty, $F(2,68) < 1$, $p = .890$, $MS_e = 684.59$. Thus, Mapping and Perceptual difficulty had additive effects in both age groups. This results, together with the absence of a Mapping × Perceptual difficulty interaction, supports the idea that in the paradigm used, the Perceptual difficulty and Mapping manipulations affected different stages of processing. With respect to the processing model outlined in the introduction, this indicates that the Perceptual difficulty manipulation merely introduced a delay with respect to the start of the 'cognitive' response selection processes.

Comparison of identical element-level rules between different rule sets. To decompose the response time cost of the arbitrary mapping, trials in the arbitrary condition were divided according to whether the element-level rule was an 'inverse' rule or not. Two repeated (nonorthogonal) contrasts tested whether (a) the Mapping effect (inverse vs. arbitrary) was significant for the comparison of identical (inverse) element-level rules, and (b) inverse vs. diagonal element-level rules within the arbitrary mapping differed in response times. Results clearly show that the set effect dominates. Reaction times to elements with identical rules in the inverse and arbitrary mapping conditions were largely different, $F(1,34) = 430.90$, $p < .001$, $MS_e = 71889.89$, while within the arbitrary mapping condition, response times were equal for inverse and diagonal rules, $F(1,34) < 1$.

This result resembles findings by Duncan (1977), who compared pure compatible and inverse mappings with a set consisting of mixed element-level rules. In the pure conditions, four serially arranged stimuli were mapped to the same serial arrangement of response keys, or the mapping was mirrored so that the leftmost stimulus was mapped onto the rightmost response etc. In mixed ensembles, a compatible mapping was used for two stimuli and a mirrored mapping for the other two. Duncan observed a mixing cost for both types of element-level mappings, meaning that response times in the mixed conditions were higher than in the pure conditions. The present result extends this finding to the comparison of inverse with arbitrary mappings.

Repetitions of element-level rule application. In a further analysis, trials were classified according to whether responses repeated or changed between trials (within a mapping block). In the former case, the same element-level rule can be applied, while in the latter the rule has to be switched within a given rule set. For compatible and inverse mappings, response repetitions and changes mainly differ in the amount of residual activation available in response-related premotor and motor systems. For arbitrary mappings however, a switch at the element-level might require deactivating the former rule as well as retrieving the current rule from long-term memory. Conversely, repeated application of the same rule might provide a shortcut to response selection, because the rule is already in the focus of working memory. The response Repetition factor is thus expected to interact with Mapping, yielding a large repetition effect only with arbitrary mappings. If episodic retrieval of a rule is a critical factor determining age differences, then the interaction of Age and Mapping should be much reduced in the case of response repetitions.

These ideas were tested with a $2 \times 3 \times 2 \times 2$ repeated measures ANOVA, including the same factors as above, plus the additional within-subjects factor of response Repetition. In addition to a main effect of response Repetition, $F(1,34) = 56.43$, $p < .001$, $MS_e = 45064.34$, the factor interacted with Age, $F(1,34) = 5.98$, $p = .20$, $MS_e = 45064.34$, and with Mapping, $F(2,68) = 56.16$, $p < .001$, $MS_e = 36198.68$.

The Repetition effect was larger for old adults, and much larger for arbitrary than for compatible or inverse mappings—in fact, single comparisons show that (a) the Repetition effect was equivalent for the latter ($F < 1$), and (b) there was no age difference in the Repetition (priming) effect for the comparison of compatible and inverse mappings ($p = 193$ for the Age-by-Repetition interaction in an ANOVA limited to these mappings).

Furthermore, a triple interaction of Age, Mapping, and response Repetitions was observed, $F(2,68) = 6.05$, $p = .004$, $MS_e = 36194.68$. Age effects in the arbitrary mapping condition were much larger when the rule changed at the element-level than when it repeated. To further investigate this interaction, separate ANOVAs were performed for rule-change and rule-repetition trials. On rule-change trials, there was a strong interaction of Age and Mapping, $F(2.68) = 12.54, p < .001, MS_e = 69248.11$. On repetition trials, the interaction was still observed, $F(2,68) = 8.94$, $p < .001$, $MS_e = 8347.15$, however, it was much weaker in magnitude: the average Age effect between the inverse and arbitrary mapping was 370 ms on change and 100 ms on repetition trials.

Log reaction times. Results from the analysis of log reaction times replicated the results from the untransformed reaction time analysis with two exceptions: first, in log reaction times, the interaction of Mapping and Perceptual difficulty was significant, $F(2,68) = 35.96, p < .001$, $MS_e = 5.27\text{e-}4$, due to a smaller difficulty effect with the arbitrary mapping (than in the other two mappings, between which the perceptual difficulty effect did not differ). Thus the proportional Perceptual difficulty effect decreases with increasingly difficult Mapping demands. This indicates that the Perceptual difficulty manipulation was relatively independent of the Mapping manipulation, since the former does not increase by the same proportion as the latter when one moves from simpler rules to the arbitrary rules. Again, this suggests relative independence of perceptual and response selection stages.

Second, the interaction of Age and Perceptual difficulty vanished in the analysis of log reaction times, $F(1,34) = 2.63$, $p = .114$, $MS_e = 2.10\text{e-}3$, suggesting that age differences in the Perceptual difficulty effect can be accounted for by proportional slowing. The interaction of Age and Mapping, however, remains significant, $F(2,68) = 7.88$, $p = .001$, $MS_e = 9.57\text{e-}3$), due to old adults' over-proportionally long response times in the arbitrary conditions. Again, no three-way interaction was observed.

In an additional analysis of log reaction times taking response repetitions into account, the response Repetition factor interacted with Mapping and with Perceptual difficulty. The effect of the arbitrary mapping was much smaller on response repetition trials than on response change trials, and the proportion that Perceptual difficulty contributed to total reaction time was actually larger on response repetition trials (this is because the same net difficulty effect was obtained in change and repetition trials: in untransformed reaction times, the interaction is not significant). More importantly, there was a three-way interaction of Age, Mapping, and Repetition, $F(2,68) = 3.55$, $p = .034$, $MS_e = 8.27\text{e-}3$. Detailed analyses of log reaction times show that Age and Mapping interacted only on change ($F(2,68) = 9.09, p < .001$, $MS_e = 1.43\text{e-}2$), and not on response repetition trials ($F(2,68) = 1.41$, $p = .25$, $MS_e = 1.02\text{e-}2$).

Errors. Overall error rate was very low at 1.8%. Only two effects reached significance in the analysis of arcsin-transformed error rates. First, the main effect of Mapping, $F(2,68) = 82.36, p < .001$, $MS_e = 2.02\text{e-}2$, was due to a much higher error rate with the arbitrary mapping (4.8%) than with the compatible (0.3%) or inverse mappings (0.4%). Single comparisons show that error rates in the latter two conditions did not differ. Second, Age and Perceptual difficulty interacted, $F(1,34) = 6.48$, $p = .016$, $MS_e = 3.29\text{e-}3$. This interaction is due to the fact that old adults made more errors in the perceptually easy condition, while young adults, as expected, had a higher error rate in the perceptually difficult condition. Closer examination reveals that the expected order of old adults' perceptual difficulty effects in errors was reversed only in the arbitrary mapping.[34]

---

[34] However, the interaction of Mapping, Perceptual difficulty, and Age was not significant, $F(2,68) = 2.46$, $p = .093$, $MS_e = 3.71\text{e-}3$. Further inspection of the error pattern

Despite these shortcomings, we can conclude from the error analysis that the interaction of Mapping and Age that was observed in the reaction time analysis can not be explained by a speed-accuracy trade-off. Although the interaction of Mapping and Age was not significant in the error analysis ($p = .131$), old adults, who were over-proportionally slowed, also made (numerically) more errors than young adults in the arbitrary mapping conditions (5.8% vs. 3.8%, old vs. young).

Discussion. Both the manipulations of early and of late difficulty were successful. Additionally, both factors interacted with age, leading to larger age effects in the more difficult conditions. First and foremost, the interaction of Age and Mapping clearly indicates that the size of the age effect obtained in the present, fairly simple reaction time task mainly depends on the degree to which retrieval and application of arbitrary rules were required—in other words, on the degree of (working) memory involvement. Results of the analyses of log reaction times and of the repetition effect further suggest a working memory locus of age differences in the current paradigm.[35] The interaction of Perceptual difficulty and Age was smaller, and might indicate age differences at an early stage of processing.

Importantly, the effect of the Perceptual difficulty manipulation was additive with regard to the interaction of Mapping and Age. In other words, the critical three-way interaction of Age, Perceptual difficulty and Mapping was absent. Although the marginal interaction of Perceptual difficulty, and Mapping—Perceptual difficulty effects tended to be larger the more time was spent in response selection—indicates that in the current paradigm, stimulus classification was not always completed before response selection started, the degree of cascaded processing was probably quite weak, so that with respect to age differences, the interaction of the two critical factors was almost perfectly additive.

Finally, comparison of the inverse with (a) the compatible and (b) the arbitrary mappings shows that an inhibitory deficit theory cannot account for age effects in the S-R compatibility paradigm. While both inverse and arbitrary mappings require inhibition of the dominant reaction towards the target, large age effects were only obtained with the arbitrary mapping. The compatibility effect in the comparison of compatible and inverse mapping rules did not interact with age, hence, inhibition of the dominant response alone cannot be the cause of age differences. This does not mean that the inhibitory deficit view of aging is flawed. The result just helps narrowing down the scope of the explanatory power of the inhibition construct with respect to age difference. Results motivate the speculation that inhibiting a response is not sensitive to aging, as long as inhibition does not take place in working memory.

To summarize, in these experiments no unequivocal support for the inhibitory deficit account was found. However, I also found no support for our hypothesis predicting that effects of perceptual difficulty should be amplified by age effects in the system representing arbitrary task rules. In unpublished data (Felbrich, 2000; Mayr & Laubrock, 2002), we obtained similar results when orthogonally varying perceptual and response selection difficulty in other experimental paradigms such as semantic categorization or visual search. This leads me to conclude that one of the key results from the additive factors literature (Sternberg, 1969, 1998), namely the relative independence of perceptual and response selection stages, is transferable to cognitive aging: with regard to the origin of age differences, perceptual and (cognitive) response selection stages appear to be independent. Apparently, the locus of age differences elicited by arbitrary task rules is at the response translation or selection stage, i.e., in a system that becomes effective relatively late in the stream of processing, after perceptual classification. Although age-related slowing affects both stages, its effects are more severe at the stage of response selection.

Modulation of age differences in Stroop interference by arbitrary task rules

Results from Experiment 1 show that age effects in an SRC paradigm are higher with arbitrary than with compatible mappings, which can be interpreted as further evidence for an age-related deficit in episodic memory. However, this age-related arbitrariness cost did not enhance the effects of earlier, perceptual difficulty manipulations, which runs counter to the predictions of the reliability of mental sets hypothesis in cascaded processing mode. Instead, results from Experiment 1are consistent with model predictions in serial mode.

One reason for the result could have been the relative distinctiveness of perceptual, cognitive (recognition and response selection), and response execution stages according to Sternberg's (1969) additive factors theory. There is ample evidence from the additive factors program that manipulations affecting early stages of the perceptual system do not usually interact with systems operating later in the chain of information processing, at least in simple 'feed-forward' tasks (e.g., Miller & Anbar, 1981; Sternberg, 1969; Shwartz, Pomerantz, & Egeth, 1977; Biederman & Kaplan, 1970). Relevant for the present case are results showing that SRC manipulations affect the cognitive and response execution stages and are thus located on a later stage than stimulus en-

---

suggests that the reversal of the perceptual difficulty effect for old adults was caused by just on participant, who had a dramatically high error rate (16%) in the perceptually easy and an average error rate (5%) in the perceptually difficult arbitrary mapping condition. I thus tend to consider the interaction of Age and Perceptual difficulty spurious. In general, error analysis seems to be dangerous with very low error rates like we observed here.

[35] This idea is also supported by results from a preliminary study (N=11 per age group) we performed using a very similar paradigm, but a smaller set size of four element-level rules. The pattern of results was qualitatively very similar, however, the size of the Age × Mapping interaction was smaller, and not significant in the analysis of log reaction times.

coding, which is affected by perceptual manipulations, such as stimulus quality or stimulus discriminability.

If this reasoning is correct, then the cascaded mode predictions of the reliability of mental sets hypothesis should be tested with difficulty manipulations acting later in the stream of processing. Ideally they would operate on a stage that is immediately used as input to the 'mental set'. In the following experiments, a manipulation of congruency of distractor and target dimension in Stroop-like stimuli was used to vary 'early' difficulty. If the mental set specifies arbitrary input-output relations in a given task, and if old adults have problems in establishing and maintaining a mental set, then age differences in the Stroop effect are expected to covary with the degree to which the task requires maintenance of a mental set. Thus large age differences in the Stroop effect are expected when an arbitrary stimulus-response mapping is required, and small age differences are expected when a compatible mapping is required.

The following series of experiments thus uses the Stroop task to manipulate cognitive difficulty at the input side of the hypothesized episodic accumulator module. The diagnostic pattern is a critical over-additive three-way interaction of Age, Stroop condition, and Mapping.

Experiment 2

Stroop task with vocal vs. manual responses and integrated vs. separated stimuli

In Experiment 2, the compatibility of the mapping of target concepts to responses was varied to manipulate the arbitrariness of task rules, and Stroop congruency was chosen as an early difficulty manipulation. Mapping compatibility was manipulated by requiring vocal and manual responses to Stroop color stimuli. The association of colors to vocal responses is compatible at the conceptual level, while the association of colors to manual responses is arbitrary: Naming a color does not have to be learned (by adults), while selecting a keypress in response to a color requires encoding of new associations. Note that at this (concept-response) level of compatibility, the association strengths of both target (color) and distractor (word) with the response are very similar, because the semantic concept activated by either is a color. A side effect of the choice of mapping manipulation is that the strength of the direct route from word distractor to response is different in the two mapping conditions. With vocal responding, word distractors have direct access to the response at a sub-semantic level, via the direct grapheme-to-phoneme conversion route, while this route is not available with manual responses. Hence, an incongruent Stroop stimulus leads to stronger priming of the wrong response in the vocal response condition. Consequently, larger Stroop effects are expected in this condition,

The main question addressed by Experiment 2 is whether age-differences in Stroop interference are larger with an arbitrary concept-response mapping, i.e. with manual responses, than with a compatible mapping, i.e. with vocal responses. A second question is related to the attentional system that contributes to possible age differences. As was discussed above, in the color-block version of the Stroop task, filtering based on the spatial location of the distractor is possible. Spatial filtering is presumable carried out by an early, posterior attentional system. On the other hand, filtering of the distractor at semantic and post-semantic stages in the standard, integrated version of the task is probably performed by a late, anterior attentional system. There is good evidence that early selection based on visuospatial attention is less affected by aging than later, anterior selection responsible for filtering the stream of cognition at a higher level. Thus, smaller age effects are expected in the separated task. Furthermore, if implementation and maintenance of the arbitrary mapping rules requires the anterior attentional system, then the degree of modulation of age differences by the mapping manipulation is expected to be smaller in the color-block task than in the integrated color-word task.

Method

Participants. Twenty-four younger (age $M = 19.7$ years, $range = 17-27$) and 24 older adults ($M = 69.5$, $range = 66-73$) participated in this experiment. Age groups did not differ in years of formal education (young, $M = 12.3$, $SD = 1.5$; old, $M = 12.9$, $SD = 3.5$), $t(46) < 1$. Age groups showed the usual pattern regarding mental speed and semantic knowledge: Young adults outperformed old adults in the digit-symbol substitution test (young, $M = 62.0$, $SD = 8.7$; old, $M = 48.3$, $SD = 7.7$), $t(46) = 5.79$, $p < .001$, while old adults had slightly better vocabulary knowledge, as measured by the MWT-A (young, $M = 30.4$, $SD = 2.6$; old, $M = 32.1$, $SD = 1.8$), $t(46) = 2.60$, $p = .013$. All subjects were healthy according to a self-rating and had normal or corrected to normal vision (Snellen accuracy $\geq$ 2/3).

Design. The experiment used a $2 \times 2 \times 3 \times 2$ design, involving the orthogonal factors of Age, stimulus-response Mapping, Stroop condition, and Task, where Task refers to whether color and word dimension were integrated on the same perceptual object, or separated such that a color block was presented above or below an achromatic word. As described in the participants section, two age groups were compared. Stimulus-response mapping was varied by using two response modalities differing in episodic demands: manual responding, with an arbitrary mapping of colors to key locations, and vocal responding, with a more compatible, albeit not 'automatic' mapping of the stimulus quality color to the response. Stroop condition (stimulus-stimulus congruency) was varied in three levels, with one third of trials belonging to congruent, incongruent, or neutral conditions each. Ink color targets were paired with color word distractors in the congruent and incongruent conditions, while in the neutral condition distractors were German adjectives. The task was always to respond to the color dimension, and to ignore the distracting word. The word and color dimensions could be integrated or separated, corresponding to a standard or a ‘color-block’ Stroop task (Hartley, 1993), respectively.

Apparatus and Stimuli. Stimuli were presented inside a rectangular white frame with a border-width of 1 pixel, subtending a visual angle of 3.5 (h) × 3.0 (v) degrees visual angle (°VA) when viewed from a distance of 70 cm. The frame was visible at all times during the experimental session and served as a fixation aid. The background of the display was dark, and the border of the stimulus frame white. The typeface used for stimulus words was Geneva, and the font size 36 points. Stimulus words were presented in lowercase letters, subtending 2.4–3.2 °VA horizontally and 1.5 °VA vertically. The color-block in the separated task was a colored rectangle with an extension of 3.5 (h) × 1.5 (v) °VA. Red, green, blue and yellow (RGB values {255,0,0},{0,255,0},{0,0,255}, and {255,255,0}) were used as the target colors. For distractors, the corresponding four German color words (`rot`, `grün`, `blau`, `gelb`) and four neutral words (`rauh`, `fern`, `dick`, `süß`, i.e., rough, far, thick, sweet) were used[36]. Congruent, incongruent, and neutral conditions were created by combining the color with the matching color-term (e.g., `rot` in red ink), one of the three non-matching color-terms (e.g., `blau` in red ink), or a neutral word (e.g., `fern` in red ink), respectively. In both the integrated and the separated task, the word could appear randomly above or below the vertical center of the screen. In the integrated task, the word was colored, i.e. both the target and the distracting information appeared on the same object at the same location. In the separated task, the irrelevant word appeared in that half of the fixation frame that was not occupied by the color block. In the manual condition, responses were registered on a regular Apple ADB keyboard and then coded for reaction times and errors by the software. In the vocal condition, a voice key (CMU Button Box) triggered by response onset was used for the registration of reaction times, and an experimenter coded errors on a prepared coding sheet, using different codes for true errors and other errors, including equipment failures due to insensitivity or oversensitivity and failures due to noise (e.g. mumbling).

Procedure. Participants were tested in two sessions, with response conditions varying between sessions. In the manual condition responses consisted of keypresses, with a mapping of stimulus colors to the keys ’<’, ’Y’, ’.’, ’-’ on a German computer keyboard (corresponding keys on an American/English keyboard are ’‘’, ’Z’, ’.’ and ’/’). Subjects were instructed to respond with their left and right index and middle fingers, with mapping of colors to keys balanced within age groups across subjects. In the verbal condition subjects had to name the colors, and responses were registered by triggering a voice-key. Each session started with 64 training trials to get used to the task, and, in the arbitrary condition, to acquire the color-to-key mapping. The irrelevant dimension was omitted during training trials, and stimuli were colored rectangles, randomly occupying the upper or lower half of the centrally presented fixation frame. Order of presentation of colors was random, and each color was presented 16 times during the training phase. After the block of training trials, participants were told that the stimuli would change, and they were instructed to respond as quickly and accurately as possible to the color, and to ignore the word meaning.

Within each session, there were two blocks of 432 trials each, further subdivided into six sub-blocks of 72 trials, giving subjects the opportunity to pause between sub-blocks. Performance feedback (mean RT, % error per sub-block) was given after each sub-block, and subjects were encouraged to reduce their error rate if it exceeded 5%. To make subjects familiar with the upcoming task, each block of 432 trials was preceded by 12 randomly selected trials with the same combination of response mode and spatial unity as the trials in the upcoming block. Spatial unity of stimuli (Task) was varied between blocks within each session. The position of the word was determined randomly from trial to trial by drawing without replacement, so that in each design cell it appeared equally often in the upper and in the lower part of the frame. In integrated-task blocks it was written in colored ink, and in the separated-task blocks it was written in white ink and accompanied by a colored block in the opposite half of the frame. Stroop condition and target color were varied within a block by drawing without replacement. Thus overall, Mapping was varied between sessions, and Task between blocks within a session, while Stroop condition and all other aspects were randomly chosen from trial to trial. Order of Mapping and of Task was counterbalanced across subjects. Half of the subjects started with manual responses, i.e. with the arbitrary mapping session, and the other half started with vocal responses, i.e. with the compatible mapping. Within each session, half of the subjects started with the integrated task, and the other half with the separated task. The same balancing order was used for the two age groups. A trial started with presentation of the empty central frame, which served as preparation cue and fixation pattern. After a fixed interval of 1000 ms the stimulus appeared inside the box. A trial was response terminated, and the next trial followed immediately, signaled by the removal of the previous stimulus.

Results

Reaction Times. After removal of outliers (1.3 %), data were analyzed using repeated measures ANOVA with Age as between-subjects factor, and Mapping/response modality (compatible/vocal vs. arbitrary/manual), Task (integrated vs. separated) and Stroop condition (congruent, neutral, incongruent) as within-subjects factors. A summary of mean reaction times and errors broken up by experimental condition can be found in Table 1.

---

[36] At the time of testing, German orthography had not been reformed, thus “rauh”, which has subsequently been replaced by “rau” was still valid.

Table 1

Means and standard errors [ms] for reaction time (columns 1-8), and mean error percentages (columns 9-12) in Experiment 2, broken up by Task, Age, Mapping/response modality, and Stroop condition.

		mean reaction time (s.e.) [*ms*]				*percent errors*			
		integrated task		separated task		integrated		separated	
		young	old	young	old	young	old	young	old
compatible/ vocal	congruent	600 (18.4)	618 (19.5)	552 (16.1)	576 (16.6)	0.2	0.3	0.3	0.1
	neutral	669 (19.0)	708 (23.3)	609 (17.7)	640 (19.7)	0.9	0.8	1.2	1.1
	incongruent	738 (23.6)	798 (28.9)	648 (20.4)	684 (23.0)	4.9	4.6	4.3	3.6
arbitrary/ manual	congruent	567 (10.4)	716 (17.6)	553 (9.7)	702 (19.8)	3.5	2.6	2.8	2.6
	neutral	582 (12.3)	737 (17.2)	564 (9.7)	725 (22.8)	3.2	2.2	3.5	2.5
	incongruent	625 (14.0)	846 (23.3)	578 (11.9)	764 (24.5)	3.9	3.1	3.2	2.6

Main Effects and 2-way interactions. Across all conditions, old adults reacted slower than young adults ($Ms = 710$ vs. 607 ms, respectively), $F(1,46) = 20.58$, $p < .001$, $MS_e = 24502.96$, leading to a global age effect of 103 ms. If we use the old-young ratio of 710/607=1.17 as a proxy for the slope of the Brinley plot, then the overall slowing factor in the task at hand appears to be rather small.

Globally, performance in the manual/arbitrary and the vocal/compatible response conditions ($Ms = 663$ vs. 653 ms) did not differ, as indicated by the insignificant Mapping main effect, $F(1,46) < 1$, $p = .357$, $MS_e = 16101.04$. However, Mapping interacted with Age, $F(1,46) = 40.91$, $p < .001$, $MS_e = 16101.04$. Age-related slowing was much larger in the manual/arbitrary than in the vocal/compatible response condition . Young adults responded faster on manual than on vocal trials ($Ms = 578$ vs. 636 ms, respectively), while old adults responded slower on manual than on vocal trials ($Ms = 748$ vs. 671 ms, respectively). This interaction was predicted by the reliability of mental sets hypothesis and replicated results from experiment 1, using a quite different paradigm. If the concept-response mapping is compatible, old adults' performance decrease is less than normally observed. In fact, if only vocal trials are analyzed, the age main effect is not significant, $F(1,46) = 1.59$, $p = .213$, $MS_e = 18311.08$.

Significant effects of Stroop condition were observed in the expected direction, $F(2,92) = 250.13$, $p < .001$, $MS_e = 1923.42$. Responses on congruent trials ($M = 610$ ms) were faster than on neutral trials ($M = 654\,ms$), $F(1,46) = 204.87$, $p < .001$, $MS_e = 1809.18$, which in turn were faster than responses on incongruent trials ($M = 710$ ms), $F(1,46) = 182.43$, $p < .001$, $MS_e = 3293.10$, yielding a facilitation effect of 44 ms, and an interference effect of 56 ms[37]. The sizes of these effects are fairly standard. Stroop condition net effects were higher for old adults (Facilitation: 5 ms; Interference: 70 ms) than for young adults (Facilitation: 38 ms; Interference: 41 ms), as indicated by the significant interaction of Age and Stroop condition, $F(2,92) = 11.01$, $p < .001$, $MS_e = 1923.42$. An inspection of the Stroop contrasts reveals that the interaction is only marginal for facilitation, $F(1,46) = 3.54$, $p = .066$, $MS_e = 1809.18$, while it is highly significant for interference, $F(1,46) = 12.43$, $p = .001$, $MS_e = 3293.10$. The two mapping conditions gave rise to different congruency effects, as indexed by the significant Mapping × Stroop condition interaction, $F(2,92) = 72.21$, $p < .001$, $MS_e = 736.05$, which was due to higher facilitation in the verbal/compatible than in the manual/arbitrary condition (70 ms vs. 18 ms facilitation for voice vs. key[38], $F(1,46) = 115.95$, $p < .001$, $MS_e = 1114.63$ for the facilitation contrast by Mapping interaction). Aggregated over Age groups, interference effects were slightly, but significantly modulated by Mapping (70 ms vs. 61 ms interference for compatible vs. arbitrary Mapping, $F(1,46) = 4.72$, $p = .035$, $MS_e = 985.09$ for the interference contrast by Mapping interaction). The small size of this effect is surprising, because a sizeable interaction had been expected based on consistent results of lower interference effects with keyboard responding reported in the literature. However, the reason for the small size of the interaction is that it is qualified by a significant triple interaction of Age, Mapping and Stroop condition reported below.

The separated color-block task (633 ms) was easier than the integrated color-word task (684 ms), as indicated by the significant main effect of Task, $F(1,46) = 87.02$, $p < .001$, $MS_e = 4281.91$. The size of the Task effect was similar for both age groups (Task by Age n.s., $F(1,46) < 1$, $p = .401$, $MS_e = 4281.91$). The size of the Stroop effect differed between the two tasks, as indicated by the Task × Stroop condition interaction, $F(2,92) = 73.24$, $p < .001$, $MS_e = 534.23$. Specifi-

---

[37] In analyses of factors involving more than two steps, reported results for the overall effect of a factor (here: Stroop condition) involve orthogonal contrasts, while reported results for further comparisons may involve nonorthogonal contrasts. In particular, evaluation of Stroop interference and facilitation effects makes use of repeated contrasts, which are nonorthogonal.

[38] The 'manual' 18 ms facilitation effect was still highly significant, $F(1,46) = 53.91$, $p < .001$, $MS_e = 570.41$. It did not differ between Age groups, $F(1,46) = 3.54$, $p = 0.66$, $MS_e = 570.41$ or Tasks, $F(1,46) < 1$, $p = .789$, $MS_e = 707.08$.

cally, both facilitation (49 vs. 38 ms, $F(1,46) = 8.69$, $p = .005$, $MS_e = 609.92$) and especially interference (78 vs. 35 ms, $F(1,46) = 65.59$, $p < .001$, $MS_e = 1375.35$) effects were larger in the integrated than in the separated task. Interestingly, Task and Mapping interacted, $F(1,46) = 20.73$, $p < .001$, $MS_e = 2690.46$. Vocal responses were faster than manual responses in the color-block version, while no such difference occurred in the integrated version of the task. However, this is probably due to a complex pattern of other interactions, in particular several interactions involving Age and one or more of the factors of Mapping, Task, and Stroop condition (see below).

Results reported thus far were fairly standard. In both response modalities the standard Stroop pattern of interference and facilitation was obtained. Modulation of interference patterns were consistent with previous reports summarized in MacLeod (1991): old adults produced more interference than young adults, interference was larger in the integrated task than in the color-block task, and interference was larger with vocal than with manual responses.

Higher-order interactions. The interaction of main theoretical importance is the triple interaction of Age, Mapping, and Stroop condition (see Table 1), $F(2,92) = 5.15$, $p = .008$, $MS_e = 736.05$. An inspection of the Stroop contrasts show that this interaction is significant for the interference contrast, $F(1,46) = 13.17$, $p = .001$, $MS_e = 985.09$, whereas it is not significant for the facilitation contrast, $F(1,46) < 1$, $p = .604$, $MS_e = 998.45$. As shown in Figure 10, old adults show a numerically larger interference effect with the manual-arbitrary than with the vocal-compatible mapping (74 ms vs. 67 ms, respectively), whereas the pattern is reversed for young adults (28 ms vs. 54 ms, respectively). Separate analyses for each age group show that for old adults, interference was equivalent between response modes, as indicated by the insignificant interaction of Mapping × interference, $F(1,23) < 1$, $p = .360$, $MS_e = 1463.23$), while young adults show the pattern often reported in the literature, of significantly less interference with keyboard responding, $F(1,23) = 21.47$, $p < .001$, $MS_e = 772.48$. Furthermore, if only vocal/compatible trials are analyzed, now again including both age groups, then there is no age difference in the Stroop effect, $F(2,92) = 2.43$, $p = .094$, $MS_e = 1760.59$, and in particular, the interaction of Age with the interference contrast is not significant, $F(1,46) = 1.86$, $p = .179$, $MS_e = 2095.53$. On the other hand, for manual/arbitrary trials both the interactions of Age with Stroop condition, $F(2,92) = 23.02$, $p < .001$, $MS_e = 989.88$, and with the interference contrast, $F(1,46) = 22.91$, $p < .001$, $MS_e = 2182.65$, are highly significant. Taken together, the pattern of results corresponds well to the predictions of the episodic buffer model. Additionally, it explains the above-reported near absence of a mapping-by-interference interaction if age is not considered. As soon as age is taken into account, the standard pattern emerges for young adults, while a qualitatively different pattern is obtained for old adults.

A second triple interaction involving Age, Stroop condition and Task also reached significance, $F(2,92) = 7.09$, $p = .001$, $MS_e = 534.23$. Age effects in interference were larger in the integrated than in the separated task, $F(1,46) = 6.97$, $p = .011$, $MS_e = 1375.35$, while there was no such effect in the facilitation contrast, $F < 1$. This possibly indicates that old adults have less of a problem ignoring irrelevant information if it can be filtered based on location. There is some evidence that location-based filtering can be achieved by early (posterior) attentional systems. In the present experiment, due to the locational uncertainty of the target above or below the vertical meridian at a stimulus-onset asynchrony of 0 ms, location-based filtering might not have been fully effective. Nevertheless, the interaction seems to indicate that as long as early attentional systems can effectively be engaged, relatively small age effects in interference can be expected. Several additional observations further support this reasoning. First, although the four-way interaction of Age, Task, Mapping, and Stroop condition failed to reach significance, $F(2,92) = 1.50$, $p = .229$, $MS_e = 503.22$, the age difference in Stroop interference was particularly large if the mapping was arbitrary, and stimulus and distractor were presented on the same object. This is indicated by results from separate analyses restricted to single Tasks. The interaction of Age × Mapping × Stroop in the interference contrast is highly significant if only trials from integrated color-word trials are analyzed, $F(1,46) = 11.00$, $p = .002$, $MS_e = 884.75$, while it is only marginal if the analyses is restricted to the color-block version of the task, $F(1,46) = 4.04$, $p = .050$, $MS_e = 648.16$. Although numerically, an increase of age effects in Stroop interference by arbitrary task rules was observed with both integrated and separated stimuli, this increase was much larger with integrated stimulus ensembles.

Finally, the triple interaction of Mapping, Stroop condition, and Task was marginally significant, $F(2,92) = 3.03$, $p = .053$, $MS_e = 335.92$ (facilitation: $F(1,46) = 7.80$, $p = .008$, $MS_e = 503.92$; interference: $F(1,46) = 2.24$, $p = .14$, $MS_e = 760.72$). While facilitation was not moderated by Task when responding manually, in the vocal response condition there was more facilitation with integrated than with separated stimuli.

Proportional reaction time measures. To take general slowing into account, two approaches were followed. First, the logarithm of reaction times was taken before aggregation to evaluate results in proportional measurement space. Second, proportional congruence scores were calculated from aggregate mean reaction times to compensate for differences in baseline reaction time.

In the analysis of log reaction times following the same scheme as above, the pattern of significant effects did not change much. However, one change affects the critical interaction of Age, Mapping, and Stroop condition, which was only marginal in log reaction times, $F(2,92) = 2.92$, $p = .059$, $MS_e = 1.36\text{e-}3$. In my opinion, this does not invalidate the above-reported re-

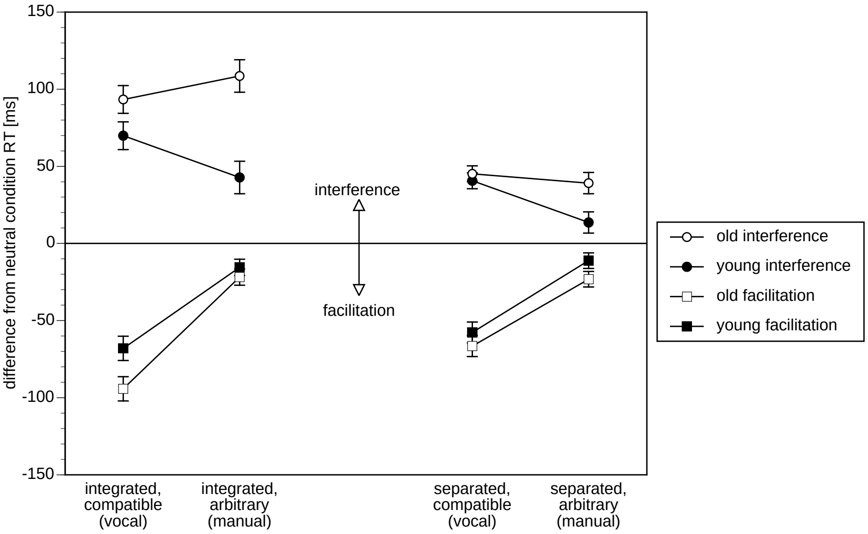


Figure 10. Experiment 2: Interference and Facilitation effects for young and old adults, as a function of Mapping and Task. Young adults are far less affected by irrelevant word distractors with manual than with vocal responding, which replicates a well-known effect. In contrast, for old adults, irrelevant word distractors influence manual responses as strongly as vocal responses. Presumably, this age-differential effect predicted by the episodic buffer hypothesis is caused by the arbitrary mapping of color concepts to keys in the manual response condition. The age difference is particularly pronounced when target and distractor are presented at the same spatial location (or integrated on the same object).

sults. While overall, the interaction effect may have failed to appear over-proportional, two additional analyses indicate that at least for a subset of contrasts, over-proportional effects were observed. First, the interference contrast of the Age×Mapping×Stroop condition interaction remains significant in the log reaction time analysis, $F(1,46) = 8.03$, $p = .007$, $MS_e = 1.71\text{e-3}$. Second, if the analysis was restricted to integrated stimuli, the critical triple interaction was significant even in log reaction times, $F(2,92) = 3.32$, $p = .041$, $MS_e = 1.12\text{e-3}$. Two further changes compared to the raw reaction time analysis were observed. First, the interaction of Age and Stroop condition, although again significant, was no longer significant for the facilitation contrast, $F(1,46) = 1.79$, $p = .188$, $MS_e = 3.33\text{e-3}$. Second, the four-way interaction of Age, Mapping, Task, and Stroop condition was not even remotely significant in the log reaction time analysis, $F(2,92) = 1.40$, $p = .252$, $MS_e = 5.75\text{e-4}$.

The second approach to compensate for general slowing used proportional congruence effects, which were calculated from individual subjects' aggregated cell mean reaction times by relating interference and facilitation to the reaction time in the neutral condition, (incongruent-neutral)/neutral and (neutral-congruent)/neutral. Since proportional *difference* scores are analyzed, an Age main effect means that Age interacts with Stroop condition in raw scores. Similarly, an Age×Mapping interaction corresponds to a triple interaction of Age, Mapping, and Stroop condition in raw reaction times. Again, the overall pattern of significance did not change much. Significant effects in the analysis of proportional *facilitation* scores were Mapping, $F(1,46) = 160.27$, $p < .001$, $MS_e = 1.84\text{e-3}$, and Task, $F(1,46) = 4.77$, $p = .034$, $MS_e = 1.27\text{e-3}$. The difference in the facilitation effect due to Mapping was larger than could be expected by the Mapping difference in the neutral response times[39]. Similarly, the facilitation effect in the integrated task was over-proportionally larger than in the separated task. No age effects in proportional facilitation were observed. The most important results from proportional *interference* scores is the fact that Age and Mapping interacted, $F(1,46) = 7.09$, $p = .011$, $MS_e = 2.28\text{e-3}$, i.e. the theoretically important interaction survives corrections for baseline response times. Further significant effects in the analysis of proportional interference scores include the main effects of Age, $F(1,46) = 8.76$, $p = .005$, $MS_e = 5.93\text{e-3}$, Task, $F(1,46) = 58.90$, $p < .001$, $MS_e = 2.94\text{e-3}$, and Mapping, $F(1,46) = 4.62$, $p = .037$, $MS_e = 2.28\text{e-3}$, suggesting over-proportional increase of the corresponding effects in the incongruent as compared to the neutral condition. However, like in the analysis of mean reaction times, the interaction of Age, Task, and Mapping was not significant.

Overall, results from the proportional analyses suggest that the age differences in facilitation observed in mean reaction times were an artifact of general slowing, which however cannot account for the other interactions observed. In particular, the interaction of Age and Mapping in the interference measure seems to be larger than expected by general slowing.

Errors. Since the overall error rate was low (2.4%), arcsin-transformed errors were analyzed using a 2(Age)

---

[39] Since 'neutral' distractors were *words*, it is not clear whether the term facilitation is really appropriate. For example, it might as well be that 'neutral' words cause some interference in preparation of vocal responses, a fact that would explain the difference in 'facilitation' between mappings.

× 2(Mapping) × 2(Task) × 3(Stroop condition) repeated measures analysis of variance with Age as between subjects factor. The error rate was not significantly different between age groups (young vs. old *Ms*: 2.6 vs. 2.1%), $F(1,46) = 1.43$, $p = .238$, $MS_e =$ 2.27e-2. Only four effects reached significance in the error analysis. Keyboard responding led to a higher error rate than vocal responding (3.0 vs. 1.8%, respectively), $F(1,46) = 21.49$, $p < .001$, $MS_e = 1.51\text{e-}2$, presumably due to the arbitrary mapping of colors to keys. Stroop condition had a large effect on error rates, $F(2,92) = 80.79$, $p < .001$, $MS_e = 1.07\text{e-}2$, which were higher in the incongruent (3.7%) than in the neutral (1.9%) or congruent (1.6%) condition. Thus errors were mainly due to incongruent distractors ($F(1,46) = 13.06$, $p = .001$, $MS_e = 1.084\text{e-}2$ for the facilitation contrast; $F(1,46) = 67.25$, $p < .001$, $MS_e = 2.87\text{e-}2$ for the interference contrast).

Congruence effects on error rate were restricted to vocal responding, as indicated by the interaction of Mapping and Stroop condition, $F(2,92) = 58.66$, $p <$ .001, $MS_e = 1.24\text{e-}2$. This effect was visible in both facilitation and interference. In fact, with manual responding, Stroop condition had no effect at all, $F(2,92) < 1$, $p = .583$, $MS_e = 8.93\text{e-}3$, while with vocal responding both facilitation, $F(1,46) = 37.95$, $p < .001$, $MS_e =$ 9.20e-3, and interference, $F(1,46) = 106.57$, $p < .001$, $MS_e = 3.13\text{e-}2$, were observed. It is hardly surprising that error rates in congruent vocal trials are very close to zero, because erroneous responding based on the word distractor cannot be detected by the experimenter. More interesting is the observation that error rates are higher with vocal (4.20%) than with manual responding (3.18%) if target and distractor are incongruent. Seemingly the relatively high degree of code overlap between distractor and response increases the likelihood of errors. While overall error rates were lower with vocal responding, in combination with incongruent stimuli the highest error rate of any condition were produced.

The last significant effect in the error analysis was the interaction of Task and Stroop condition, $F(2,92) =$ 6.23, $p = .003$, $MS_e = 5.81\text{e-}3$. The Stroop effect was somewhat larger when stimulus and distractor were spatially integrated.[40]

Age did not interact with any of the factors. Age differences in error rate were not differentially affected by Mapping, $F(1,46) < 1$, $p = .376$, $MS_e =$ 1.51e-2, Stroop condition, $F(2,92) < 1$, $p = .495$, $MS_e = 1.07\text{e-}2$, or Task, $F(1,46) < 1$, $p = .479$, $MS_e =$ 7.48e-3, nor did age modulate any of the interactions. In particular, the critical triple interaction of Age, Stroop condition, and Mapping was far from significant, $F(2,92) < 1$, $p = .807$, $MS_e = 5.19\text{e-}3$. Thus the critical results from the reaction time analysis can not be explained away by a speed-accuracy tradeoff.

Discussion

Most aspects of the pattern of results were consistent with predictions of the reliability of mental sets hypothesis. First, slowing was larger when the target-response mapping was arbitrary. This fits the assumption that with compatible target-response mappings, reliability of mental sets plays no role because sets are not needed. Second, the critical Age × Task × Stroop condition interaction was observed: age-differences in interference were amplified by an arbitrary target-response mapping. The costs introduced by an arbitrary concept-response mapping seem to outweigh the benefits from the concomitantly less-strong influence of the distractor via the direct route (i.e. the less compatible distractor-response mapping), which for young adults leads to a reduction in interference. Thus, results support the idea that mental sets specifying arbitrary associations between internal stimulus representations and responses become less reliable in old age.

The fact that the triple interaction is weaker in the color-block version than in the standard, integrated version of the task hints at a likely locus of the age deficit at a system implementing late attentional control. Age deficits get smaller when early filtering based on spatial location is possible. Taken together, results are fairly consistent with the model sketched in the introduction. This encouraged us to further explore the limits of the observed effects in the following experiments.

Two results deserve special discussion. First, there is the fact that facilitation was as high as interference with vocal responding, but much smaller than interference with manual responding. The fact that facilitation was much larger with vocal responding is consistent with a dual-route conception: under conditions of high distractor-response compatibility (i.e. with vocal responses), the congruent word primes the response via the direct grapheme-to-phoneme route, so that relatively little additional input via the controlled, semantic route is sufficient to evoke a response.

In the extreme case, responses on a proportion of trials might have been based entirely on the wrong dimension. On congruent trials there is no way for the experimenter to detect this error. Because word-based responses are identified as errors in both the neutral and the incongruent conditions, such goal-neglect may actually lead to an increase in the facilitation effect. On vocal trials, because of the high degree of distractor-response compatibility, these erroneous responses should be particularly fast. Inclusion of undetectable word-response errors would thus lead to an apparently large facilitation effect. On the other hand, with keyboard responding, there is no priming of the response along the direct route, instead, a translation of concept to response is required for both target and distractor. Even if word reading should lead to faster conceptual activation than color perception, this could only lead to a priming of the stimulus-response rule, not to responding based directly on the distractor. Indirect evidence supporting this interpretation would thus come from incongruent color naming *error* trials, where reaction times should be much faster than on correct

---

[40] In the current experiment, 'spatial' integration is confounded with 'object-based' integration (see Duncan, 1984)).

trials if responses are based on word reading, i.e., if goal-neglect is a contributing factor. This is exactly what was observed: On incongruent trials in the vocal response condition, where inadvertent reading leads to an erroneous response, error trial RTs (mean±s.e. young: 649±15 ms, N=310; old: 638±14 ms, N=271) were much faster than correct RTs (young: 717±3.8 ms, N=6652; old: 745±3.3 ms, N=6641), whereas on neutral trials there was not much of a difference between error and correct response times (error vs. correct young 696±31 ms, N=74 vs. 661±3; old 674±25, N=63 vs. 678±3 ms). With keyboard responding, this pattern was not observed. To the contrary, errors were somewhat slower (young: 642±15 ms, N=243; old: 1108±67 ms, N=189) than correct responses (young: 621 ms, old: 834 ms) on incongruent trials—however, standard errors were large indicating a less consistent pattern. Since RTs on incongruent error trials in the vocal condition were in the range of congruent RTs, one could by extrapolation assume that a proportion of very fast trials in the congruent condition were also based on reading rather than color naming.

The second factor responsible for the differential size of facilitation effects between response modalities is probably the choice of the neutral condition. Recall that the distractor in the neutral condition was a proper word with an associated lexical entry. It is likely that the neutral condition is not truly neutral, but caused some interference at the lexical level, which might have a larger influence on verbal than on manual responding due to the greater association of verbal responses to lexical activation. Possibly, the choice of neutral condition made facilitation effects in the vocal response condition appear larger and interference effects smaller than if a nonword or an unpronounceable letter string had been chosen.

The second point that needs to be discussed is the extent to which slowing was larger in the manual/arbitrary than in the vocal/compatible response condition. In fact, there was hardly any slowing with vocal responding in the congruent condition. One aspect that could possibly shed light on the reason for the large Age×Mapping effect is the counterintuitive result that young adults responded faster with an arbitrary mapping (on manual trials) than with a compatible mapping (on vocal trials). This effect was particularly pronounced in the neutral and incongruent conditions (vocal-manual difference of 16, 66, and 92 ms for congruent, neutral, and incongruent conditions, young adults), and was somewhat unexpected, since the S-R mapping was arbitrary in the manual condition. Except for the difference in the congruent condition[41], this can probably largely be explained by the fact that as a by-product of the change in target-response compatibility, distractor-response compatibility also changed. The direct grapheme-to-phoneme route is only available with vocal responses, because a word distractor has much larger overlap with the internal code used for preparing a naming response than with the manual response code. Thus manual responses could be faster than vocal responses simply because the influence of a word distractor on the internal (spatial?) code used to prepare a manual response is weaker. The fact that for young adults, the manual advantage correlates with Stroop interference renders this explanation likely. If the explanation is correct, then the corollary decrease in distractor-response compatibility outweighs any negative effects of increased working memory demands due to an arbitrary S-R mapping. Old adults responded slower on manual than on vocal trials. This effect was enhanced because of higher facilitation on compatible-vocal trials (vocal-manual difference of -112, -57, and -64 ms for congruent, neutral, and incongruent conditions, old adults), which might partly be explained by the fact that congruent vocal responses are a mixture of correct and error trials, see above. Seemingly for old adults, the decrease in target concept-response compatibility, presumably necessitating the involvement of mental sets, was sufficient to outweigh the potential benefits of reduced interference due to weaker distractor stimulus-response compatibility.[42] The different contributions of stimulus-response compatibility at a conceptual and a sub-semantic level of Stroop interference are further investigated in a later chapter in Experiment 5.

Let me again summarize the main results of the present experiment: While facilitation effects were similar for old and young adults, and while young adults produced the well-known pattern of smaller interference effects with manual than with vocal responding, old adults could not profit from the potentially beneficial effects of reduced priming of the wrong response when responding manually. The age effect in interference was much higher with manual than with vocal responding. Apparently, the arbitrary mapping from color concept to manual response amplified age differences in the Stroop effect. This was particularly obvious in the condition

---

[41] I do not want to delve into the reasons for the Mapping effect in the comparison of young adults' response times for the congruent conditions. In principle, there could be a manual response modality advantage for young adults, which then should also show up in simple reaction times. In fact, results from a comparison of simple reaction times in the vocal and manual modalities suggest that manual responses are somewhat faster. For example, in a comparison of response modalities, Nebes (1978) reported a mean reaction time of 242 ms for manual simple RT, and of 321 ms for vocal simple RT. However, in the current context this simple explanation is rendered unlikely by the observed reaction time equivalence for manual and vocal responses in the congruent conditions, $F(1,23) = 1.19$, $p = .286$, $MS_e = 5089.72$. Instead, there is some evidence for a strategic effect. Relatively strong inhibition of the word distractor in the vocal version of the integrated task appears to be carried over from neutral and incongruent to the congruent trials. While for young adults on congruent trials, vocal and manual reaction times are identical in the separated task, vocal RTs are significantly slower than manual RTs in the integrated task. The interaction of Mapping and Task in the analysis restricted to young adults on congruent trials is highly significant, $F(1,23) = 14.41$, $p < .001$, $MS_e = 502.29$.

[42] Again, an explanation based only on the differential speed of motor systems seems unlikely, in particular because Nebes (1978) observed an advantage of manual as compared to vocal simple reaction times not only for young adults, but also for older adults (manual RT old: 278 ms, vocal RT old: 339 ms).

with integrated stimuli, where filtering by an early attentional system was less likely.

Age effects were considerably larger with manual than with vocal responding even in the neutral condition. The present experiment did not include a simple reaction time control condition, hence it cannot be firmly concluded that this effect was caused by the arbitrary mapping, and not by the fact that different motor systems were differentially affected by age. A more cautious interpretation of the results of Experiment 2 is that age effects in Stroop interference depend on output characteristics of the task.

Experiment 3

Stroop task with two vocal response conditions

Results from Experiment 2 were compatible with prediction derived from the hypothesis that the representation of mental sets becomes less reliable in old age. However, we cannot yet firmly conclude that the arbitrary S-R rules introduced by manual responding are responsible for this effect. It might well be that the manipulation of response modality itself was the critical factor. In fact, some independent evidence suggests the possibility that age-related slowing may depend on the response modality. For example, in a study that investigated age differences between vocal and manual response systems using simple reaction time tasks (Nebes, 1978), no age differences were found in vocal responses (average age difference: 18 ms), while at the same time a significant age difference was obtained in the manual modality (average age difference: 36 ms). Thus even in a simple reaction time task, i.e., in a situation where cognitive demands are nearly absent, an interaction of age group and response modality was obtained. Thus two follow-up experiments were designed to address the confounding between response modality and arbitrariness of concept-to-response mapping that was present in Experiment 2. In both Experiments 3 and 4 a vocal response condition with an arbitrary mapping of verbal labels to color concepts was compared with a standard Stroop color-naming task. Experiment 4 used a different set of verbal labels than Experiment 3, and also included a manual response condition, thus providing one contrast for effects of mapping arbitrariness, and one contrast for effects of response modality within arbitrary mappings.

In Experiment 3, all responses were given vocally in a Stroop-like task. Two stimulus-response mappings were used. One response condition replicated the compatible vocal condition from Experiment 2, while the second response condition introduced an arbitrary S-R mapping that required participants to 'name' colors by pronouncing male first names, which had been associated with the four target colors in an initial learning phase. Only integrated stimuli were used, since these had caused the largest age effects in the previous experiment.

There are similarities and differences in the way arbitrary mappings were used in the current experiment and in the studies performed by MacLeod and Dunbar (1988) to illustrate the development of automaticity in the Stroop task. Common to both approaches is the fact that arbitrary relations between stimuli and responses were learned. Whereas in the current experiment, the arbitrary mapping was defined by an association of color stimuli to male name responses, MacLeod and Dunbar's subjects learned to associate four arbitrary shape stimuli with color responses. Thus in MacLeod and Dunbar's study the association of stimuli to color concepts was arbitrary, while the (color) concept-response association was pre-learned. In contrast, in the present experiment, associations of stimuli to color concepts were prelearned, while (color) concept-response set associations were arbitrary. With an arbitrary response set instead of an arbitrary stimulus set, the semantic activation of the color concepts should be as fast as with the compatible set. Thus, it is likely that any effect of the retrieval of a mapping rule can only begin after there is sufficient activation of a color concept. The arbitrary-vocal condition therefore conceptually resembles a condition with keyboard responding.

Method

Participants. Sixteen young (age $M = 20.6$, $range = 19 - 22$ years) and 16 old adults ($M = 74.8$, $range = 68 - 81$) participated in the experiment. All subjects were recruited from the University of Potsdam psychology department subject pool and were paid for their participation. Age groups did not differ in the total years of formal education including higher education (young, $M = 13.0$, $SD = 0.7$; old, $M = 13.1$, $SD = 3.9$), $t(30) < 1$. Young adults outperformed old adults in the Digit Symbol Substitution test (young, $M = 62.3$, $SD = 7.1$; old, $M = 47.4$, $SD = 8.9$), $t(30) = 5.24$, $p < .001$. Old adults performed marginally better than young adults in the MWT-A vocabulary test (young, $M = 31.4$, $SD = 1.6$; old, $M = 32.4$, $SD = 1.5$), $t(30) = 1.80$, $p = .082$. All participants were healthy according to a self-rating and had normal or corrected to normal vision.

Design, Stimuli and Procedure. The design was a $2 \times 2 \times 3$ mixed factorial with age as between-subjects factor and the within-subjects factors of concept-response Mapping (compatible vs. arbitrary) and Stroop condition (congruent, neutral, and incongruent). Stimuli were standard Stroop color-word stimuli identical to the ones used in the integrated color-word task in Experiment 2. The compatible mapping condition replicated the vocal response condition of Experiment 2, i.e., color names had to be pronounced. The response in the arbitrary mapping condition consisted in pronouncing the German male first names 'Horst', 'Bert', 'Jan', and 'Kurt', which had been associated with colors within the experimental context. At the level of individual elements, the mapping of colors to names was counterbalanced between subjects and matched between groups defined by age and response-order. The experiment was run in two blocks, with Mapping varying between blocks. Order of blocks was counterbalanced between subjects and matched between age groups. In a learning phase immediately before the start of the arbitrary block, colors were associated with male first names. The learning phase was further subdivided. First, there was a block of 32 trials (8 per color) during which the color-to-name mapping was visible near the bottom of the screen (by presenting a row of four colored bars below a row of the corresponding to-be-associated names). This block could be repeated if subjects felt they needed more practice. Second, there was a further block of 32 trials during which the mapping was not visible. During the whole learning phase, the stimulus was a colored bar (subtending 3.7 °VA vertical and 1.8 °VA horizontal) without a distracting word, and the task consisted of pronouncing the associated name. To make the two blocks as similar as possible, the training phase was also included in the compatible mapping condition, where the color of the bar had to be named, and color names were presented above the color patches near the bottom of the screen.[43]

After the learning phase, the experiment proper started, which consisted of 432 Stroop color-word trials per mapping. Within a mapping block, Stroop condition was randomly chosen with replacement, such that words, which were congruent, neutral, and incongruent with respect to the word color, were presented on one third of the trials each.

The task consisted of reacting to word color, either by naming it or by pronouncing the associated male first name. In the incongruent condition, each of the four target colors red, green, blue, and yellow was equally often paired with words signifying the other three response-set colors. In the neutral condition, the words `süß`, `kalt`, `dick`, and `fern` (sweet, cold, thick, far) were equally often paired with each target color.

A trial started with the presentation of a screen-centered empty 'fixation frame' that surrounded the position of the to-be-presented word. During the learning phase, this frame was filled with a color after 1000 ms. During the experimental phase, a colored word was presented inside the frame after 1000 ms. The fixation frame subtended 3.7 °VA horizontally and 4.3 °VA vertically. The stimulus words subtended 2.6 to 3.4 °VA horizontally and 1.8 °VA vertically. An inter-trial interval of 1000 ms followed detection of a response by the voice-key. Because a small error rate was expected based on the results of the previous experiment, instructions emphasized speed. However, it was mentioned that correct responding was also important.

Results

Reaction times. Reaction times from correct trials (that were not immediately preceded by an error trial) were filtered to remove outliers according to the criteria described in the general methods section, with a minimum RT threshold of 200 ms and a maximum RT threshold of 2500 ms. Applying these criteria led to a removal of 2.1 % of trials (2.3% young, 1.9% old adults).

Mean reaction times (see Table 2) were subjected to a 2×2×3 repeated measures ANOVA with Age as between subjects factor, and Mapping and Stroop condition as within subjects factors. All three main effects were significant, Age, $F(1,30) = 19.17$, $p < .001$, $MS_e = 9189.85$, Mapping, $F(1,30) = 95.72$, $p < .001$, $MS_e = 3919.28$, and Stroop condition, $F(2,60) = 105.24$, $p < .001$, $MS_e = 1790.14$. Old adults were slower than young adults, responding with first names (i.e. in the arbitrary mapping condition) was slower than responding with color names, and both Stroop facilitation, $F(1,30) = 95.67$, $p < .001$, $MS_e = 1511.95$, and Stroop interference, $F(1,30) = 75.63$, $p < .001$, $MS_e = 3118.12$, were observed. Age interacted with Stroop condition, $F(2,60) = 10.27$, $p < .001$, $MS_e = 1790.14$: compared to young adults, old adults produced larger interference effects, $F(1,30) = 13.26$, $p = .001$, $MS_e = 3118.12$, but similar facilitation effects, $F(1,30) = 1.91$, $p = .18$, $MS_e = 1511.95$. In contrast to the results of Experiment 2 (and of Experiment 1), Age and Mapping did not interact significantly, $F(1,30) < 1$. One reason for the lack of an interaction might be that the manual response modality was responsible for this particular interaction in Experiment 2. However, alternative explanations are conceivable, as will briefly be discussed at the end of this section.

There was a significant interaction of Mapping and Stroop condition, $F(2,60) = 35.57$, $p < .001$, $MS_e = 634.98$. Single comparisons reveal that this effect was particularly pronounced in facilitation, $F(1,30) = 38.15$, $p < .001$, $MS_e = 1024.14$, which was much smaller in the arbitrary than in the compatible condition, but also significant in interference, $F(1,30) = 6.08$, $p = .020$, $MS_e = 1438.11$. Both components of the Stroop effect were larger with compatible color responses than with arbitrary name responses. One is tempted to speculate that the difference in facilitation might be caused by the fact that in the case of compatible mappings, confirmation along the semantic route is fast enough to arrive at the response buffer before distractor-caused priming along the direct route has started to decay. The fact that the size of interference was only slightly reduced with name responses could indicate that Stroop interference has quite a large lexical component.

Results are indecisive regarding the critical interaction of Age, Mapping, and Stroop condition, possibly due to the small number of participants combined with a relatively large variance in responding with arbitrary names. On the one hand, the interaction is far from significant, $F(2,60) = 1.82$, $p = .17$, $MS_e = 634.98$. On the other hand the numerical trend was in the expected direction, and the effect was even larger if the Stroop effect, i.e. the difference in reaction time between the incongruent and congruent Stroop conditions, was analyzed. In the analysis of the difference measure, the interaction of age and mapping was marginal, $F(1,30) = 3.46$, $p = .073$, $MS_e = 1316.96$. Furthermore, while there was a large Stroop effect in both age groups with the compatible mapping (*Ms* = 112 *vs.* 169 *ms*, young vs. old), the Stroop effect was relatively small for young

---

[43] Unsurprisingly, none of the participants requested an additional training block in the compatible condition, whereas several did so in the arbitrary condition.

Table 2

Means (and standard errors) for reaction time [ms] (columns 1-4), and mean error percentages (columns 5-6) in Experiment 3, broken up by Age, Mapping, and Stroop condition.

		mean RT (s.e.) [ms]		*percent errors*	
		young	old	young	old
compatible/ (colors)	congruent	533 (9.39)	607 (18.66)	0.3	0.2
	neutral	601 (9.42)	681 (15.77)	0.8	0.7
	incongruent	653 (14.96)	773 (24.24)	4.4	2.8
arbitrary/ (names)	congruent	718 (27.70)	801 (19.20)	1.3	2.0
	neutral	731 (25.33)	833 (21.92)	1.4	1.3
	incongruent	747 (23.39)	915 (35.02)	1.6	2.0

adults with the arbitrary mapping, while it remained large (albeit reduced) for old adults ($Ms = 31$ *vs.* 121 *ms*, young vs. old). But since none of the critical effects reached significance, it seems as if the arbitrary mapping did not lead to an enhancement of the age effect in Stroop interference in this experiment. However, an analysis of repetition effects suggests that this is not the full story.

Response repetition effects. Although the effect is labeled ‘response repetition’, it should be clear that in the current paradigm, a response repetition is also a target stimulus repetition, and furthermore, in the arbitrary mapping condition, involves a repetition of the rule applied on the last trial. If the arbitrarily mapped name is still active because it was used on the previous trial, it might be able to provide a short-cut to response selection. On response repetition trials in the arbitrary mapping, a rule does not have to be retrieved from long term memory, while on change trials, the former rule has to be deactivated and a new rule has to be retrieved. Thus strong ‘cognitive’ response repetition effects can be expected in the arbitrary mapping condition. These should be larger than response repetition effects with the compatible mapping, which are mainly due to priming of the response at the output end.

This interaction of Mapping and Response repetitions was observed, $F(1,30) = 104.17, p < .001, MS_e = 8741.86$, as shown in Figure 11, in addition to a strong main effect of Response repetitions, $F(1,30) = 211.78$, $p < .001$, $MS_e = 15893.79$. Repetition effects were particularly pronounced under arbitrary mapping conditions, where response repetition trials were 285 ms faster than response change trials, compared to a response repetition effect of 90 ms with the compatible mapping. If the arbitrarily mapped name was still active, then responses were on average as fast as on response repetition trials with the compatible mapping ($Ms = 583$ *vs.* 576 *ms*, respectively). Response repetition also modulated the effects of Stroop condition, $F(2,60) = 12.33$, $p < .001$, $MS_e = 961.22$. The Stroop effect was larger on change than on repetition trials. Importantly, there was also a four-way interaction involving Response repetition, Age, Mapping, and Stroop condition, $F(2,60) = 4.91$, $p = .011$, $MS_e = 731.86$. Although this interaction is rather complex, it indicates that the response repetition factor modulates the interaction that is critical for the present argumentation. The interaction pattern basically means that the modulation of the Stroop effect by response repetitions is larger for old than for young adults with the arbitrary mapping, while it is larger for young than for old adults with the compatible mapping. Indeed, if only trials in which the response changed with respect to the previous trial are analyzed, the interaction of Age, Mapping, and Stroop condition almost reaches significance, $F(2,60) = 2.75$, $p = .072$, $MS_e = 1075.48$. Analyses using the Stroop effect difference measure show a significant interaction of Age, Mapping, and Response repetition, $F(1,30) = 13.59, p = .001, MS_e = 1024.97$, which is depicted in the right panel of Figure 11. Using this measure, Age and Mapping interact when the response changed, $F(1,30) = 5.32$, $p = .028$, $MS_e = 2059.06$, but not when the response was repeated, $F(1,30) = 2.43$, $p = .130$, $MS_e = 1547.02$.

Proportional measures. An ANOVA of the mean of the logarithm of raw reaction times produced results that resembled the untransformed ones and are therefore not reported in detail here. However, it seems that in the analysis of log reaction times the interaction of Mapping and Stroop condition, $F(2,60) = 73.05$, $p < .001$, $MS_e = 1.03\text{e-}3$, is relatively larger than in the raw reaction time analysis, while the critical interaction of Mapping, Stroop condition, and Age, which already failed to reach significance in the raw reaction time analysis, was even weaker in the analysis of log reaction time ($F(2,60) < 1$).

Similarly, if the Stroop effect difference measure is computed from log reaction times, the interaction of Age and Mapping that was marginal in the raw reaction time analysis becomes insignificant, $F(1,30) = 1.74$, $p = .197$, $MS_e = 2.08\text{e-}3$. Taking response repetitions into account, the four-way interaction of Response repetition, Age, Mapping, and Stroop condition is only marginal in the log reaction time analysis, $F(2,60) = 2.92$, $p = .062$, $MS_e = 1.25\text{e-}3$.

A second attempt to correct for baseline reaction time differences was made by using proportional scores calculated by dividing the Stroop effect, i.e.

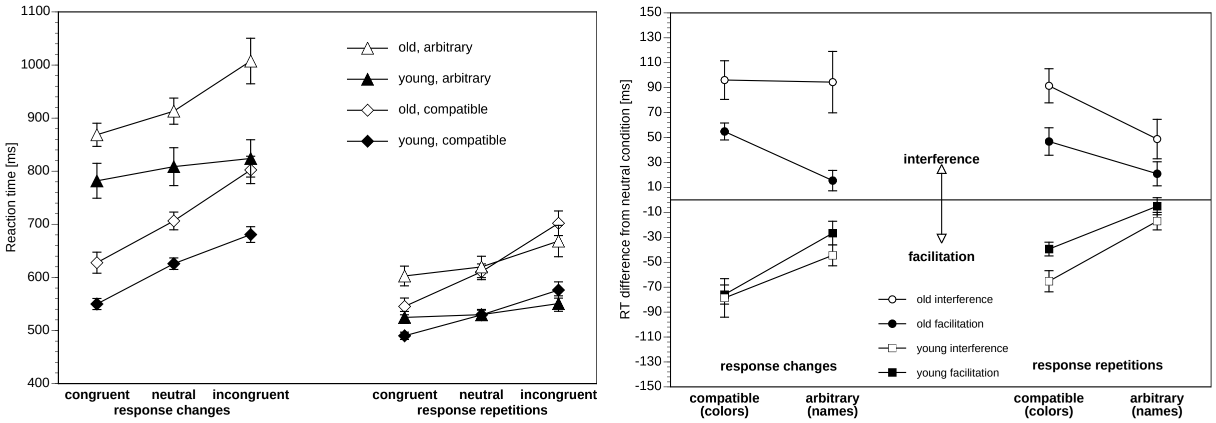


Figure 11. Experiment 3: *Left panel:* Reaction times broken up by the experimental factors and target (response) repetitions vs. changes. Repetition effects were particularly prominent in the arbitrary mapping condition. On response change trials, there was a tendency for an increased age difference in Stroop interference with the arbitrary as compared to the compatible mapping. *Right panel:* Interference and Facilitation effects as a function of Age, response Repetition, and Mapping. Large repetition effects led to a reduction of the age difference in the Stroop effect that was otherwise (i.e., on response change trials) observed in the arbitrary mapping conditions.

the incongruent-congruent reaction time difference, by the neutral condition reaction time. In this proportional measure of the Stroop effect, there was a significant three-way interaction of Age, Mapping, and the Response repetition factor, $F(1,30) = 10.205$, $p = .003$, $MS_e = 1.87\text{e-}3$. This suggests that the larger age effect in Stroop interference on arbitrary mapping, response change trials is more than expected by proportional slowing. Further significant effects in this analysis include the Age main effect, $F(1,30) = 9.92$, $p = .004$, $MS_e = 1.58\text{e-}2$, and the Mapping main effect, $F(1,30) = 113.85$, $p < .001$, $MS_e = 4.75\text{e-}3$.

Errors. About 1% of all trials were lost due to errors not obviously related to wrong responses, such as equipment error or early triggering of the voice-key. The overall error rate in the remaining trials (with reaction time outliers removed) was low (about 1.5%). A 2×2×3 ANOVA with Age, Mapping, and Stroop condition revealed a significant main effect of Stroop condition, $F(2,60) = 27.37$, $p < .001$, $MS_e = 2.37\text{e-}4$, with no facilitation ($F < 1$), but interference in the expected direction, $F(1,30) = 34.05$, $p < .001$, $MS_e = 5.13\text{e-}4$. No other main effect was significant. There were two significant two-way interactions. Mapping and Stroop condition interacted, $F(2,60) = 28.06$, $p < .001$, $MS_e = 2.37\text{e-}4$, indicating larger interference, $F(1,30) = 28.06$, $p < .001$, $MS_e = 3.59\text{e-}4$, and facilitation, $F(1,30) = 5.17$, $p = .030$, $MS_e = 2.02\text{e-}4$, with colors than with arbitrary names. In fact, the effect of Stroop condition on the error rate in the arbitrary mapping condition was insignificant, $F < 1$. Mapping and Age also interacted, $F(1,30) = 12.64$, $p < .001$, $MS_e = 2.71\text{e-}5$: old adults committed less errors with the compatible than with the arbitrary mapping, while the reverse was true for young adults. This latter result is however mainly due to the fact that young adults produced a relatively large error rate on incongruent trials in the compatible (color) response condition—in both congruent and neutral trials they actually made less errors than in the respective arbitrary mapping conditions, and the age difference was negligible (see Table 2, page 46). All other interactions failed to reach significance, although the three-way interaction of Age, Mapping, and Stroop condition only marginally did so, $F(2,60) = 2.65$, $p = .079$, $MS_e = 1.78\text{e-}4$. In this interaction, the interference contrast was significant, $F(1,30) = 4.89$, $p = .035$, $MS_e = 3.59\text{e-}4$. Young adults made more errors than old adults on incongruent trials with the compatible mapping (color responses), but less errors than old adults on incongruent trials with the arbitrary mapping (male first name responses).

Discussion

Although there was a tendency in the expected direction, the interaction of Age and Mapping in the Stroop effect failed to reach significance. However, the interaction was modulated by response repetition. If the reaction and the concept-response rule on a given trial was not primed by the name pronounced on the previous trial, the critical interaction of Age and Mapping in the Stroop effect measure was significant, but if it was primed, the modulating influence of arbitrariness vanished. Thus it appears that in the case of rule repetitions, the primed response (or the primed rule) provides a shortcut to response selection. When the rule is still activated in working memory, episodic accumulators can largely be bypassed, whereas they are needed for episodic retrieval of the rule on change trials.

Taken together, the results seem to confirm that arbitrary S-R rules lead to more interference-prone selection processes in old age. However, in comparison to Experiment 2, effects were rather weak and therefore they are not very convincing. Because the major design change between experiments was in response modality, one reason could be that old adults not only have a specific deficit in the episodic buffers, but that this is also limited to manual responding. However, it is easy to think of a number of alternative explanations for the difference in effect size. For example, memory demands by the arbitrary mapping were rather high in this experiment, because the potential set of all male first names is rather large. Other than in keyboard responding, there is no visible reminder of the response set members. Thus in addition to the arbitrary mapping rules, the response set representation itself is less well defined. Responding with male first names induces additional demands, which might have functioned like an 'age simulation' by preventing young subjects from developing an efficient memory representation of the rules that provides a direct connection between stimulus and response. Another possible explanation is that power was weak, because only 16 participants were tested per age group. The tendency for an interaction of Age, Mapping, and Stroop condition might have become significant with a larger sample. Finally, there is of course also the possibility that results of Experiment 2 were only due to sampling error.

Experiment 4

Stroop task with vocal/compatible, vocal/arbitrary, and manual/arbitrary S-R mappings

To rule out these alternative explanations, sample size in Experiment 4 was increased to 24 participants per age group, and a condition that replicated the manual/arbitrary condition of Experiment 2 was included. Furthermore, like Experiment 3, Experiment 4 also included a vocal/arbitary condition. However, in contrast to the previous experiment, a well-defined set of responses was used, namely the set of the natural numbers one to four. Representation of the vocal/arbitrary response set itself thus does not impose high memory demands, thereby more closely resembling the manual response set. Participants learned to associate the numbers to colors in the training phase of the experiment. This vocal/arbitrary condition was compared with a manual/arbitrary condition and a vocal/compatible condition.

The rationale behind this choice of response sets was to compare (a) two conditions that are similar in memory demands on a set-level, but differ in response modality (manual vs. vocal/arbitrary), and (b) a condition with low memory demands, vocal/compatible, with conditions high in memory demands, manual and vocal/arbitrary. Single comparisons also allow to compare two conditions that differ in memory demands, but use the same response modality (vocal/compatible vs. vocal/arbitrary). If the large age effect in Stroop interference observed in the manual response condition in Experiment 2 was due to the memory demands imposed by the arbitrary S-R mapping, then the age difference in the interference measure should be large in the arbitrariness contrast. If on the other hand the earlier result was due to interference proneness of manual responding in old age, then the age difference in interference should be larger with keyboard responding than with vocal responding, even if the S-R mapping is equally arbitrary in the manual and in the vocal/arbitrary response condition.

To summarize, in comparison to the Experiment 3, there were two major changes. First, a manual response condition was added. Second, the response set in the manual arbitrary condition was changed from male first names to numbers, which more closely resemble manual responses, because the extension of the response set is well-known. Due to the fact that the factor Mapping now had three levels, the number of trials per Mapping was smaller than in Experiment 3.

Method

Participants. Twenty-four young (age $M = 19.9$, $\mathit{range} = 17 - 28$ years) and 24 old adults ($M = 71.9$, $\mathit{range} = 65 - 91$) participated in the experiment. Age groups were comparable with respect to total years of formal education (young, $M = 11.9$, $\mathit{SD} = 1.9$; old, $M = 13.1$, $\mathit{SD} = 3.5$), $t(46) = 1.51$, $p = .137$. Young adults performed better than old adults on the Digit Symbol Substitution test (young, $M = 59.8$, $\mathit{SD} = 7.0$; old, $M = 48.3$, $\mathit{SD} = 7.7$ points), $t(46) = 11.54$, $p < .001$, while there was a tendency for old adults to perform better on the MWT-A vocabulary test (young, $M = 31.8$, $\mathit{SD} = 2.4$; old, $M = 32.9$, $\mathit{SD} = 1.7$ points), $t(46) = 1.77$, $p = .085$. All subjects were healthy according to a self-rating and had normal or corrected to normal vision.

Apparatus. The apparatus was the same as in the previous experiment, with one addition. While voice responses were registered using the CMU button box, as described above, manual responses were registered on the Apple keyboard connected via ADB.

Design, Stimuli and Procedure. The design was a $2 \times 3 \times 3$ mixed factorial, involving the between-subjects factors of Age group (young, old), and the within-subjects factors of Mapping (vocal/compatible, vocal/arbitray, manual/arbitary) and Stroop condition (congruent, neutral, incongruent). Stimuli and timings were identical to the ones used in Experiment 3, and so was the task in the vocal mapping conditions, with the exception that a different set of responses, namely the words EINS, ZWEI, DREI, and VIER (German for one, two, three, and four), were used in the vocal/arbitrary mapping condition. In the manual condition, the index and middle fingers of both hands rested on the ’<’, ’y’, ’.’, and ’-’ keys (corresponding to ’‘’, ’z’, ’.’, and ’/’ on a US keyboard), to which colors were mapped. Both the mapping of colors to keys and of colors to numbers were counterbalanced between subjects and matched between age groups. During training trials in the manual mapping condition, the mapping rules were displayed near the bottom of the screen in graphical format, using colored boxes arranged in the same left-to-right order as the corresponding keys. During training, the key cap label was also presented on screen, above the corresponding color boxes.

Mapping was blocked, and Stroop condition was randomized within blocks. Due to the inclusion of a manual response block while keeping the overall number of trials constant at 864, the number of trials per Mapping was reduced as compared to Experiment 3. In the current experiment, there were 288 trials per Mapping, and participants were given the opportunity to rest after every 72th trial.

Results

Reaction times. Outlier removal according to the criteria detailed in the general methods section led to a removal of .68 % of trials (.75% young, .61% old adults). Individual participants' cell means of the remaining trials were subjected to a $2\times3\times3$ repeated measures ANOVA with Age as between subjects factor.

For the Mapping factor, two orthogonal (Helmert) contrasts were used, the first comparing the effect of *arbitrariness*, i.e. vocal/compatible vs. the mean of vocal/arbitrary and manual, and the second comparing the effect of *response modality* within the arbitrary mappings. If a significant effect involving the former contrast was found, additional tests were performed using single comparisons of vocal/arbitrary and manual with the vocal/compatible condition. For the Stroop factor, nonorthogonal repeated contrasts were used to estimate the effects of interference and facilitation. The pattern of mean reaction times and errors broken up by the experimental factors is shown in Table 3.

Main Effects and two-way interactions. As expected, all three main effects were significant, Age, $F(1,46) = 14.22$, $p < .001$, $MS_e = 34473.49$, Mapping, $F(2,92) = 25.56$, $p < .001$, $MS_e = 18689.71$, and Stroop condition, $F(2,92) = 225.78$, $p < .001$, $MS_e = 1559.20$. Old adults were slower than young adults (mean age effect 117 ms), both interference, $F(1,46) = 174.81$, $p < .001$, $MS_e = 2823.09$, and facilitation, $F(1,46) = 128.92$, $p < .001$, $MS_e = 1807.04$, were reliable, and both Mapping contrasts were significant: Responses in the arbitrary mapping conditions were slower than in the compatible mapping, $F(1,46) = 16.99$, $p < .001$, $MS_e = 7251.39$, and vocally responding with numbers was slower than responding manually, $F(1,46) = 30.99$, $p < .001$, $MS_e = 15251.10$. Single comparisons show that response times in the vocal/arbitrary condition ($M = 776\,ms$) were also slower than in the vocal/compatible condition $F(1,46) = 37.14$, $p < .001$, $MS_e = 19496.02$, whereas vocal/compatible and manual response times ($Ms = 675\,vs.\,676\,ms$) did not differ from each other at this level of analysis, $F(1,46) < 1$. Results for the Mapping main effect are thus comparable with results obtained in Experiments 2 and 3 .

Age (marginally) failed to modulate the Mapping effect on an overall level, $F(2,92) = 2.94$, $p = .058$, $MS_e = 18689.71$, mainly due to the complete absence of an interaction of Age with the response modality contrast comparing manual and vocal/arbitrary (number) responses, $F(1,46) < 1$, $p = .44$, $MS_e = 15251.10$. However, the Age × Mapping interaction was significant for the arbitrariness contrast, $F(1,46) = 6.62$, $p = .013$, $MS_e = 7251.39$. In the supplemental analysis using single comparisons, the interaction was significant for the comparison of manual and vocal/compatible (color) responses, $F(1,46) = 7.81$, $p = .008$, $MS_e = 13696.49$, replicating the result of Experiment 2. Age effects also tended to be larger with the vocal/arbitrary than with the vocal/compatible mapping, $F(1,46) = 2.25$, $p = .14$, $MS_e = 19496.02$. Thus results of Experiments 2 and 3 were replicated, in particular, a substantial Age × Mapping interaction for the vocal/compatible vs. manual comparison, and a negligible Age × Mapping interaction for the comparison of compatible and arbitrary mappings within the vocal response modality. Despite the lack of a significant interaction for the latter comparison, age effects were numerically larger in the arbitrary condition. Thus overall, age effects appear to be larger with arbitrary than with compatible concept-to-response mappings.

Stroop condition and Mapping interacted, $F(4,184) = 24.64$, $p < .001$, $MS_e = 1039.75$. The pattern of interactions shows that the two arbitrary conditions did not differ in facilitation, $F < 1$, but in interference, which was larger with manual responses, $F(1,46) = 9.21$, $p = .004$, $MS_e = 2413.81$, while both facilitation and interference effects were higher in the vocal/compatible (color) than in the arbitrary conditions, $F(1,46) = 31.06$, $p < .001$, $MS_e = 6458.65$ for facilitation, $F(1,46) = 23.09$, $p < .001$, $MS_e = 2413.81$ for interference. Single comparisons for Mapping show that for both pairwise comparisons with the vocal/compatible condition, the interaction was significant (manual vs. vocal/compatible: $F(2,92) = 35.27$, $p < .001$, $MS_e = 1237.82$; vocal/arbitrary vs. vocal/compatible: $F(2,92) = 35.17$, $p < .001$, $MS_e = 12.37$), due to both higher interference and facilitation in the vocal/compatible color naming condition[44]. Taken together, results regarding the interaction of the arbitrariness contrast with Stroop condition suggest that interference and facilitation are larger with compatible than with arbitrary mappings. Surprisingly, results for the response modality contrast indicate that the Stroop effect caused by word distractors is no larger with vocal/arbitrary than with manual responses. Speculations about possible reasons will be brought forward in the discussion.

Finally, Stroop condition and Age interacted, $F(2,92) = 16.34$, $p < .001$, $MS_e = 1559.20$, with old adults experiencing more interference, $F(1,46) = 29.44$, $p < .001$, $MS_e = 941.03$, but not facilitation, $F(1,46) < 1$, $p = .553$, $MS_e = 602.347$, than young adults.

---

[44] *Interference*, manual vs. vocal/compatible: $F(1,46) = 6.78$, $p = .012$, $MS_e = 1131.73$; vocal/arbitrary vs. vocal/compatible: $F(1,46) = 31.66$, $p < .001$, $MS_e = 1750.07$; *Facilitation*: manual vs. vocal/compatible: $F(1,46) = 33.37$, $p < .001$, $MS_e = 1829.26$; vocal/arbitrary vs. vocal/compatible: $F(1,46) = 19.27$, $p < .001$, $MS_e = 2251.29$.

Table 3
Means (and standard errors) for reaction time [ms] (columns 1-4), and mean error percentages (columns 5-6) in Experiment 4, broken up by Age, Mapping, and Stroop condition.

		mean RT (s.e.) [ms] young	*mean RT (s.e.) [ms]* old	*percent errors* young	*percent errors* old
compatible/ (vocal)	congruent	563 (15.86)	640 (24.29)	0.4	0.0
	neutral	647 (21.85)	700 (19.77)	1.6	1.0
	incongruent	705 (25.80)	797 (26.28)	7.6	4.9
arbitrary/ (vocal)	congruent	698 (23.25)	793 (34.20)	4.6	2.8
	neutral	711 (22.21)	832 (39.14)	3.8	3.0
	incongruent	732 (23.24)	887 (45.50)	3.3	3.9
arbitrary/ (manual)	congruent	573 (18.41)	711 (24.63)	4.9	2.4
	neutral	602 (19.40)	725 (20.40)	5.9	1.5
	incongruent	626 (20.61)	821 (27.66)	5.8	2.7

Three-way interaction. The critical three-way interaction of Age, Mapping, and Stroop condition was significant, $F(4,184) = 3.37$, $p = .011$, $MS_e = 1039.75$.

Leaving the neutral Stroop condition aside for a moment, the interaction of Age × Stroop condition (congruent vs. incongruent) × Mapping was significant for the arbitrariness contrast, $F(1,46) = 4.85$, $p = .033$, $MS_e = 2971.84$, while the interaction was not significant for the response modality contrast within the arbitrary mappings, $F(1,46) < 1$, $p = .91$. This seems to strongly support an explanation for specific age effects that is based on memory demands rather than on output modules.

A closer look at the triple interaction pattern, now again including the neutral condition, reveals that the interaction was significant for all three pairwise comparisons between mappings (manual vs. vocal/arbitrary: $F(2,92) = 3.20$, $p = .045$, $MS_e = 984.68$, vocal-arbitrary vs. vocal/compatible: $F(2,92) = 3.23$, $p = .044$, $MS_e = 896.74$; manual vs. vocal/compatible: $F(2,92) = 3.60$, $p = .031$, $MS_e = 1237.82$). The fact that with, but not without inclusion of the neutral condition, the interaction was significant for the comparison of the two arbitrary mappings suggests that age effects in the processing of 'neutral' distractors may differ between response modalities. Further dissecting the interaction, in the comparison of manual and vocal/compatible mappings, the three-way interaction was significant for interference, $F(1,46) = 5.75$, $p = .012$, $MS_e = 1131.73$, but not facilitation $F(1,46) < 1$, $p = .62$, $MS_e = 1829.26$. This replicated results of Experiment 2. Young adults experienced far less interference in the manual response condition, while for old adults, interference in the manual and the vocal/compatible condition were about equal. Facilitation was higher in the vocal/compatible condition, and the increase in facilitation as compared to the manual condition was the same for both age groups.

In the comparison of vocal/arbitrary and vocal/compatible mappings, the interaction, which had not been significant in Experiment 3, was significant in the current experiment, $F(2,92) = 3.60$, $p = .031$, $MS_e = 1237.82$, possibly due to the larger sample size. The three-way interaction was due to age differences in facilitation, $F(1,46) = 5.66$, $p = .022$, $MS_e = 2614.57$, but not interference, $F(1,46) < 1$, $p = .7$, $MS_e = 1177.12$. As reported above, both age groups experienced higher interference in the vocal/compatible than in the vocal/arbitrary condition. However the increase did not differ between age groups, unlike the facilitation contrast: Young adults showed almost no facilitation (13 ms) in the vocal/arbitrary condition, while old adults did (39 ms). This pattern was reversed in the vocal/compatible condition, where young adults produced a larger facilitation effect (84 ms) than old adults (60 ms).

Finally, in the comparison of manual and vocal/arbitrary mappings, the three-way interaction was significant for both the interference, $F(1,46) = 7.31$, $p = .01$, $MS_e = 1206.90$, and the facilitation contrasts, $F(1,46) = 5.81$, $p = .02$, $MS_e = 1729.32$. The facilitation pattern crossed over between Mappings and Age groups, with young adults showing a higher facilitation effect when responding manually (29 vs. 14 ms, manual vs. vocal/arbitrary), and old adults showing a higher facilitation effect when responding vocally with numbers (13 ms vs. 39 ms, manual vs. vocal/arbitrary). Old adults produced generally larger interference effects than young adults, and especially so in the manual response condition.

In summary, on the one hand, and most importantly, the age difference in the Stroop effect increases when arbitrary mapping rules are involved. While the age difference is relatively small in the vocal/compatible condition (142 vs. 157 ms, old vs. young), it is sizeable in both the vocal/arbitrary (34 vs. 94 ms) and the manual (53 vs. 110 ms) conditions. On the other hand, the vocal/arbitrary and the manual/arbitrary mapping seem to affect different components of the Stroop effect in the two age groups. It is not entirely clear how to interpret this pattern, thus the state of the neutral condition has to be discussed, which I will do after presenting results from response repetition, proportional, and error analyses. Regardless of the debatable state of the neutral condition, the reaction time results show that the age difference in the Stroop effect is much larger with arbitrary than with compatible concept-to-response mappings. In contrast, there is no age difference in the Stroop effect if different response modalities are compared that both use arbitrary mapping rules.

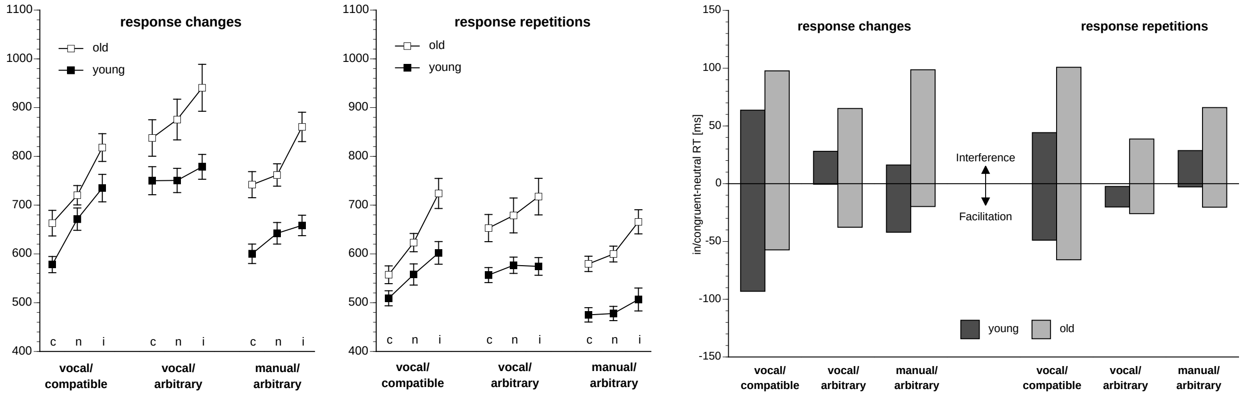
Response repetition effects. Since Experiment 3 has shown that access to the elements of arbitrary mappings is much more efficient if an element is repeated on subsequent trials, another set of analyses were performed that included the response Repetition factor. Response Repetition modulated the theoretically most important interaction, as indicated by the significant four-way interaction of Age × Mapping × Stroop condition × response Repetition, $F(4,184) = 5.87$, $p < .001$, $MS_e = 1063.22$, depicted in Figure 12. Separate analyses for each level of response Repetition show that Age, Mapping, and Stroop condition interacted on response change trials, $F(4,184) = 5.961$, $p < .001$, $MS_e = 1306.68$, and failed to interact on response repetition trials, $F(4,184) < 1$, $p = .48$, $MS_e = 1409.77$. On response change trials, this three-way interaction was significant for all three pairwise comparisons between mappings (manual vs. vocal/compatible $F(2,92) = 4.69$, $p = .012$, $MS_e = 1366.10$; vocal/arbitrary vs. vocal/compatible: $F(2,92) = 7.33$, $p = .001$, $MS_e = 1524.65$; manual vs. vocal/arbitrary: $F(2,92) = 5.62$, $p = .005$, $MS_e = 1029.30$). Like in the analyses of all trials, in the manual vs. vocal/compatible comparison, the interaction was due to the interference contrast, $F(1,46) = 7.25$, $p = .01$, $MS_e = 2446.66$ (facilitation: $F < 1$), whereas for the vocal/arbitrary vs. vocal/compatible comparison the interaction was due to the facilitation contrast, $F(1,46) = 9.09$, $p = .004$, $MS_e = 3529.66$ (interference: $F < 1$). Comparing manual and vocal/arbitrary responses, both interference, $F(1,46) = 7.39$, $p = .009$, $MS_e = 1668.13$, and facilitation, $F(1,46) = 9.80$, $p = .003$, $MS_e = 2165.32$, contributed to the three-way interaction. Here, the age difference in interference was larger in the manual than in the vocal/arbitrary condition (82 vs. 37 ms mean age difference old-young, manual vs. vocal/arbitary), and the reverse was true for the age difference in facilitation (-22 vs. 37 ms). While young adults showed larger facilitation effects than old adults when responding manually, in contrast to old adults they showed absolutely no facilitation effect when responding with numbers.

The neutral condition involved words with no obvious color association. However, these words still have lexical entries, thus, it is possible that they interfere more with a verbal coding or maintenance of mapping rules. If old adults are more affected by verbal distractors, it might be more reasonable to investigate the effect of the Stroop manipulation by leaving out the 'neutral' condition and comparing only congruent with incongruent trials, i.e. the 'Stroop effect' difference measure. For response change trials, the pattern of results suggests that there is absolutely no age difference in the Stroop effect with vocal/compatible responses, whereas old adults produce a higher Stroop effect than young adults with both manual and vocal/arbitrary responses (see Table 4). Thus it appears that in the conditions under investigation, the arbitrary mappings led to age differences in the Stroop effect.

Table 4
Experiment 4: Stroop effect on response change trials [ms].

Mapping	young	old
manual/arbitrary (keys)	58	118
vocal/arbitrary (numbers)	29	103
colors/compatible (colors)	157	155

Proportional measures. Results of the analysis of log-transformed reaction times largely paralleled the untransformed reaction time results, albeit with some differences. First, Age and Mapping interacted significantly, $F(2,92) = 5.46$, $p = .006$, $MS_e = 2.36\text{e-}2$, while the interaction was only marginal in the raw reaction time analysis. In proportional space, old adults were much slower than young adults in the manual response condition, and less so in the vocal response conditions. The age effect did not differ significantly between the two vocal response conditions. The critical interaction between Age, Mapping, and Stroop condition remained significant in the analysis of log-transformed reaction times, $F(4,184) = 3.05$, $p = .018$, $MS_e = 1.35\text{e-}3$. The interaction was significant for the comparison between manual and vocal/arbitrary responses, $F(2,92) = 4.16$, $p = .019$, $MS_e = 1.13\text{e-}3$, it was only marginally significant for the comparison between the two vocal response conditions, $F(2,92) = 3.03$, $p = .053$, $MS_e = 1.56\text{e-}3$, and there was only a tendency towards significance for the comparison between manual and vocal/compatible responding, $F(2,92) = 2.15$, $p = .12$, $MS_e = 1.36\text{e-}3$. However, if the neutral condition is left out of the analysis, the arbitrariness contrast that had been significant in the untransformed analysis was only marginally significant in the log-RT analysis, $F(1,46) = 2.89$, $p = .096$, $MS_e = 3.00\text{e-}3$. The response modality contrast comparing the arbitrary mappings was far from significant, as in the untransformed analysis. Thus the neutral condition must have had some differential influence on proportional scores. To check this, the interference and facilitation contrasts from the single comparisons between mappings were inspected. Significant age modulations of Stroop interference were found for the comparison between the manual and vocal/compatible mappings, $F(1,46) = 10.79$, $p = .002$, $MS_e = 7.77\text{e-}4$. Significant age modulations of facilitation were found for the other two comparisons, manual vs. vocal/arbitrary, $F(1,46) = 5.36$, $p = .025$, $MS_e = 1.01\text{e-}3$, and vocal/arbitrary vs. vocal/compatible, $F(1,46) = 4.90$, $p = .032$, $MS_e = 1.91\text{e-}3$. The simplest explanation for this pattern of results seems to be that the amount of interference caused by the 'neutral' word in the vocal/arbitrary condition differs between age groups.



(a) Reaction times (b) Interference and facilitation effects

Figure 12. Experiment 4: Response repetition modulates age effects in the interaction of Stroop condition and mapping. *(a)* Mean reaction times on change trials (left panel) or repetition trials (right panel), split up by Age, Mapping, and Stroop condition (c: congruent, n: neutral, i: incongruent). Filled symbols represent young adults, and open symbols old adults. *(b)* Interference and facilitation effects as difference from the neutral condition (represented by the line at $y = 0$ ms). The left set of columns shows effects on response change trials, and the right set of columns on response repetition trials. Positive values indicate interference, and negative values facilitation.

Taking response repetitions into account, the four-way interaction of the Repetition factor with Age, Mapping, and Stroop condition remained significant in the analysis of log reaction times, $F(4,184) = 4.36$, $p = .002$, $MS_e = 1.78\text{e-}3$. If the analysis was limited to response change trials, the critical interaction of Age, Mapping, and Stroop condition was significant in proportional space, $F(4,184) = 5.24$, $p = .001$, $MS_e = 1.65\text{e-}3$. Like in the analysis of untransformed scores, the effect was due to a larger age difference in the Stroop effect with the arbitrary mappings than with the compatible mapping (i.e., there was a significant interaction with the mapping contrast, $F(1,46) = 5.65$, $p = .022$, $MS_e = 4.42\text{e-}3$), whereas the age difference in the Stroop effect remained at a similar magnitude for the comparison of the two arbitrary mapping conditions (i.e., there was no interaction with the modality contrast, $F(1,46) < 1$, $p = .895$).

In analyses of the proportional Stroop effect measure, (incongruent-congruent)/neutral, both the main effects of Age, $F(1,46) = 9.48$, $p = .003$, $MS_e = 2.44\text{e-}3$, and of Mapping, $F(2,92) = 62.69$, $p < .001$, $MS_e = 4.04\text{e-}3$ were significant. Both the contrasts for arbitrariness, $F(1,46) = 106.70$, $p < .001$, $MS_e = 6.54\text{e-}3$ and for response modality, $F(1,46) = 11.06$, $p = .002$, $MS_e = 7.44\text{e-}3$, contributed to the Mapping effect. Although the interaction of Age × Mapping failed to reach significance, $F(2,92) = 2.34$, $p = .102$, $MS_e = 4.04\text{e-}3$, the arbitrariness contrast of the interaction was significant, $F(1,46) = 4.29$, $p = .044$, $MS_e = 6.54\text{e-}3$, while there was no age difference in the comparison of manual and vocal/compatible mappings, $F < 1$. In the single comparisons with the compatible mapping, the interaction of Mapping with Age approached significance for the manual, $F(1,46) = 3.88$, $p = .055$, $MS_e = 4.06\text{e-}3$ as well as for the vocal/abitrary mapping, $F(1,46) = 2.86$, $p = .098$, $MS_e = 4.34\text{e-}3$.

Taken together, the pattern of results from the proportional analyses resembles the pattern obtained in the raw RT analysis. It cannot be firmly concluded that the triple interaction is over-proportional. Nevertheless, the pattern is quite consistent, and even in the cases where effects did not reach significance in the proportional analyses, they only marginally failed to do so. Furthermore, using log-RT scores may not be the best way to correct for general slowing, and the proportional Stroop effect scores are less reliable than untransformed means.[45]

Errors. The average error rate was at a low 3.3%, thus results from the error analysis must be treated with some care. Young adults produced more errors ($M = 4.2\%$) than old adults ($M = 2.5\%$), $F(1,46) = 6.21$, $p = .016$, $MS_e = .13$. Stroop condition influenced error rates in the expected order, $F(2,92) = 29.87$, $p < .001$, $MS_e = 1.35\text{e-}2$, with least errors on congruent ($M = 2.5\%$), little more on neutral ($M = 2.8\%$), and most on incongruent trials ($M = 4.7\%$). Facilitation was not significant in errors, but interference was. In the error analysis, there was no interaction of Age and Stroop condition, $F < 1$. Thus in the current experiment larger Stroop effects for old adults were confined to reaction times, however see below for the interaction of Age, Mapping, and Stroop condition. Mapping also influenced error rates, $F(2,92) = 17.82$, $p < .001$, $MS_e = 1.44\text{e-}2$. More errors occurred with the arbitrary mappings than in the vocal/compatible con-

---

[45] In additional analyses using the regression prediction and z-transformation approaches suggested by Faust et al. (1999), the three-way interaction of Age, Mapping, and Stroop condition was highly significant for the arbitrariness contrast ($F(1,46) = 8.36, p = .006$ for z-transformed scores, $F(1,46) = 10.30, p = .002$ for regression predictions), but not significant for the response modality contrast ($F < 1$ for both measures). The pattern of results strongly supports the conclusions from the untransformed analysis, and additionally suggests that the age difference in the modulation of Stroop interference by memory demands cannot be accounted for by differences in baseline reaction time, i.e., it is over-proportional.

dition ($Ms = 3.9$ *vs.* 3.6 *vs.* 2.6% for manual vs. vocal/arbitrary vs. vocal/compatible), while the error rate did not differ between the two arbitrary mappings. Age interacted with Mapping, $F(2,92) = 10.30$, $p < .001$, $MS_e = 1.44\text{e-}2$. Detailed analyses show that there was no age difference with the vocal/arbitrary mapping, $F(1,46) = 1.03$, $p = .31$, $MS_e = 2.32\text{e-}2$, a marginal age effect with the vocal/arbitrary mapping, $F(1,46) = 3.18$, $p = .081$, $MS_e = 1.14\text{e-}2$, and a significant age effect with the manual/arbitrary mapping, $F(1,46) = 16.57$, $p < .001$, $MS_e = 1.87\text{e-}2$. The age difference in the manual response condition was quite sizeable ($Ms = 5.5$ *vs.* 2.2%, young vs. old), indicating that with keyboard responding, young adults might have traded off accuracy for speed. Stroop condition and Mapping interacted, $F(4,184) = 27.09$, $p < .001$, $MS_e = 1.32\text{e-}2$. Stroop effects were much larger in the vocal/compatible than in the arbitrary conditions, while the interaction was not significant for the response modality comparison within the two arbitrary response conditions. In fact, the Stroop effect in error rate as well as the interaction with Mapping was entirely due to the vocal/compatible condition, because neither of the two effects was significant in an analysis limited to the arbitrary mappings, both Fs< 1. The above-mentioned speed-accuracy trade-off does not affect the critical three-way interaction of Age, Mapping, and Stroop condition, $F(4,184) = 3.41$, $p = .010$, $MS_e = 1.32\text{e-}2$. In both arbitrary Mapping conditions, old adults showed a numerical increase in errors from congruent to incongruent Stroop trials, while young adults showed either a small increase (in the manual condition) or even showed a decrease in errors (in the vocal/arbitrary condition). In contrast, young adults' error rate was more affected by Stroop condition than old adults' with the vocal/compatible mapping. Although these results do not directly parallel the reaction time results, at least they show that age differences in criterion settings for the speed-accuracy trade-off cannot be the sole reason for the reaction time pattern with respect to the critical interaction. Young adults in general produce more errors, and especially do so (a) when responding manually, or (b) on incongruent trials in the vocal/compatible condition. However, in the arbitrary mapping conditions the age difference in errors is not sensitive to Stroop condition; in fact it is (numerically) reduced in the incongruent conditions.

Discussion

The most important results from the current experiment are that (a) the predicted amplification of an age difference in the Stroop effect by arbitrary mappings was obtained, and that (b) arbitrariness of the mapping rules, and not a change in response modality, was critical in modulating the interaction of Age and Stroop condition. The pattern of results replicates the pattern obtained in Experiment 2, where vocal/compatible and manual responding was compared, and gives additional credibility to the preliminary conclusions drawn from Experiment 3, where compatible and arbitrary vocal mappings were compared. Effects regarding the interaction of Age, Mapping, and Stroop condition that were marginal in Experiment 3 were significant in the present experiment, which involved a larger sample size.

Two aspects of the results will be discussed in more detail below, first, the fact that the Stroop effect was not higher with the vocal/arbitrary than with the manual mapping, and second, the problematic interpretation of interference and facilitation measures.

Mapping, response modality, and Stroop condition. One interesting aspect of the current results is the pattern obtained for the interaction of Stroop condition and Mapping. For the arbitrariness contrast, interference and facilitation were larger with the compatible than with the arbitrary mappings. While this appears to support a dual-route model of Stroop interference, interestingly, arbitrariness seems to play a larger role than response modality: For the comparison between the two arbitrary mappings, the Stroop effect was numerically larger for manual than for vocal/arbitrary responses, whereas a naive dual-route model would predict a larger Stroop effect in the vocal than in the manual condition. This is because a verbal output corresponding to the distractor is activated along the grapheme-to-phoneme route, and a verbal output has a greater chance to interfere with vocal than with manual responses. The dual-route models can be saved if it is assumed that output priming by a distractor decays with time. A similar assumption is made in recent Stroop models to explain negative SOA effects (Roelofs, 2003; Stafford, 2003). Because responses in the vocal/arbitrary condition take on average 100 ms longer than in the manual condition, more decay will have taken place in the former at the point in time the activation along the episodic route reaches response selection.

Stroop facilitation and interference. The interpretation of the results regarding the critical interaction of Age, Mapping, and Stroop condition is somewhat complicated by the fact that the 'neutral' Stroop condition had different effects in different mapping conditions for the two age groups. Thus a less clear pattern of results was obtained with regard to facilitation and interference effects. In most conditions (if response repetitions are not taken into account), facilitation and interference effects were more or less symmetric. The exception are old adults in the manual response condition, for whom the incongruent distractor caused a large cost if compared to the neutral word distractor, while the congruent word caused only a very small benefit.[46]

---

[46] Separate analyses of the Stroop effect were performed for each Age group and Mapping, using a polynomial contrast for the Stroop effect. While the linear contrast was significant in all analyses, the quadratic term was not significant in either age group for vocal/arbitrary responses, it was small, but significant in either age group for vocal/compatible responses (with a different sign of the curvature between age groups because of young adults' untypical results of larger facilitation than interference effects), it was not significant for young adults in the manual response condition, and it was highly significant for old adults in the manual response condition.

If we compare these results with results of the previous experiments, there are commonalities and differences. First, for the comparison of manual and vocal/compatible responses, like in Experiment 2, old adults were much more susceptible to Stroop *interference* than young adults when responding was manual than when the standard, compatible/vocal response format was used. Also like in Experiment 2, there was no age difference in the interaction of Stroop *facilitation*. However, unlike in Experiment 2 young adults' facilitation effects in the manual response condition of the current experiment were similar in magnitude to their interference effects (in Experiment 2, interference effects were substantially larger than facilitation effects). Second, for the comparison of the two vocal mappings, unlike in Experiment 3, the interaction of Age, Mapping and Stroop condition was mainly due to differences in *facilitation*. Like in Experiment 3, facilitation effects in the vocal/arbitrary condition were larger for old adults than for young adults. Unlike in both Experiments 2 and 3 facilitation effects were larger for young adults than for old adults in the vocal/compatible condition. Thus for the comparison of facilitation between the vocal mappings, there was a crossover interaction of Mapping with Age.

In summary, at the level of the Stroop effect results nicely replicate between experiments, whereas they partly fail to replicate at the level of its components, namely facilitation and interference. This could be related to the fact that these component measures are less reliable than the Stroop effect. Although a number of studies in the literature have addressed the contribution of a variety of cognitive processes to the component measures, there does not appear to be consensus on how to interpret these relative effects. Furthermore, even results that appear to be fairly consistent are not necessarily generalizable across experimental settings. For example, while facilitation effects are typically smaller than interference effects in a standard Stroop task (MacLeod, 1991), this was not the case in the paradigm used here. Instead, in the current experiment as well as in Experiments 2 and 3, interference and facilitation were pretty much symmetric (especially for young adults) in the vocal/compatible condition. I can only speculate about the reasons for this, which might include the relatively larger number of trials in the current experiments, or the particular choice of neutral word distractors.

What could be responsible for the lack of reliability of the interference and facilitation measures? Two points likely contribute: First, responding on congruent trials is a mixture of goal neglect and true facilitation. Second, the neutral distractor itself might cause some interference.

The fact that facilitation is likely a mixture of answers based on the distractor (word) and answers based on the target dimension (color) has recently been emphasized (e.g., MacLeod, 1998; MacLeod & MacDonald, 2000; De Jong et al., 1999; Dunbar & MacLeod, 1984). This inadvertent reading or goal neglect explanation is particularly, but not exclusively, applicable in the vocal/compatible condition, where reading of a congruent distractor leads to a very fast response. Responding based on word reading is an error, because task instructions are not followed. The error is undetectable for the experimenter, because the same word is pronounced, regardless of whether the response is based on target or distractor. Even with arbitrary mappings, source confusion on some trials might lead to a response based on the distractor. With arbitrary, unlike with compatible mappings, distractor-based responding does not lead to a large reaction time benefit, because responding cannot be based on automatic activation alone, but the matching response has to be retrieved.[47] However, if the speed of activation of the mapping rule differs between word and color stimuli, goal neglect could cause a loss of reliability of facilitation scores.

One further reason for the lack of reliability of the interference and facilitation scores could be the fact that the 'neutral' word itself elicits some interference. In fact, several component processes contribute to interference caused by neutral words, and the degree to which they contribute might differ between participants and mapping conditions. There is evidence that 'Stroop interference' as the difference in reaction time between incongruent and neutral conditions is sensible for the type of neutral condition. For example, when responding is vocal, interference is largest if a letter string is used as the neutral distractor, and progressively gets larger with neutral conditions consisting of neutral words, color-related words, and color-words that are not in the response set. This has been interpreted as suggesting that Stroop interference consists of a lexical (neutral word-letter string), a semantic relatedness (color-related vs. neutral word), a semantic relevance (color-related vs. color nonresponse), and a response set membership component (response set member vs. not). However, it is not clear whether these effects of the neutral condition are also relevant when responding is manual. To my knowledge, only one study compared manual and vocal responses in the Stroop task and at the same time included a number of different neutral conditions to explicitly investigate the different contributions of lexical, semantic, and response set membership effects (Sharma & McKenna, 1998). These authors found that all of the mentioned component processes contribute to interference with vocal responding, while only response set membership is relevant with manual responding.[48] Unfortunately, Sharma and McKenna did not include a congruent condition, so that it is impossible to tell whether facilitation would still be observed with respect to a neutral condition consisting of unpronounceable letter

---

[47] There will still be some benefit, because responding on congruent trials will profit from a lack of interference at the level of retrieval of the mapping rule.

[48] However, this argument is based entirely on significance of pairwise comparisons. Interestingly, even with manual responding, interference effects were numerically affected by the type of neutral condition in the same order as mentioned above.

strings (e.g., XXXX), in comparison to which interference effects were largest.

To conclude, the interaction of Age, Mapping, and Stroop condition followed the pattern predicted by the hypothesis of a reduced reliability of mental sets in old age. Whether the interaction was mainly observed in facilitation or in interference differed between experiments and mapping conditions. One possible reason for this is that the component measures might not be very reliable. However, further work with several neutral conditions is needed to more rigorously determine which components of the Stroop effect are influenced to what degree by which psychological processes. In retrospect, it is unfortunate that I used only a single neutral condition, because additional neutral conditions, e.g. using pronounceable and unpronounceable nonwords, could have shed some light on the underlying mechanisms.

Experiment 5

Spatial Stroop task with response modality and mapping completely crossed

Although the results of Experiment 4 suggest that the memory demands of a reaction time task are critical in modulating age differences, this was really only shown for the comparison with a vocal/compatible condition. To firmly conclude that response modality does not affect the critical interaction, it seems desirable to include a manual/compatible condition, especially because independent evidence suggests that response modality itself might be a critical modulator of age differences (Nebes, 1978; Doose & Feyereisen, 2001). Furthermore, there is evidence that the anterior cingulate cortex (ACC), which is a system that has consistently been found to be involved in Stroop interference resolution, is organized in a somatotopical fashion. For example, Turken and Swick (1999) found that performance in a spatial Stroop task of a patient with a focal right hemisphere ACC lesion depended on the response modality used. Under the same task requirements, she was impaired when giving manual responses, but not vocal responses. Importantly, the patient did not show a deficit in executive control. Thus our current results could possibly be reinterpreted to indicate that a modality-specific impairment in a system associated with response selection is a critical mediator of age differences in performance. This would constitute a different mechanism than the episodic buffer model proposed in the introduction, which was assumed to have a more central, executive locus and should therefore not be as sensitive to changes in response modality. To exclude this alternative, it seems advisable to include a manual/compatible mapping condition.

Since manual responses have an inherent spatial component due to the response key layout, stimuli with a spatial semantic are needed to obtain conceptually compatible relations. I therefore chose a spatial Stroop paradigm, which allows for the orthogonal manipulation of conceptual compatibility and response modality. Target stimuli in the compatible mapping conditions were arrows and words. It was assumed that for both left-pointing arrows and the word LEFT there were relatively reliable pre-experimental associations with the spatial concept of 'leftness'. Responses in the manual response condition consisted of left and right keypresses. For vocal responses, the response set consisted of the words 'left' and 'right', thus tapping the same directional semantics as manual responses. These compatible mappings were compared with arbitrary mapping conditions, which mapped stimulus shape or stimulus color as target dimensions onto the same left/right response sets used in the compatible conditions.

A second problem regarding the conclusions to be drawn from the previous experiments is related to a confounding of concept-response mapping and distractor-response compatibility. Recall that according to dual-route models, a word distractor primes a phonological output module even before the stimulus is semantically processed. Because vocal responses use a phonological output representation, interference caused by word distractors is thus expected to be particularly high with vocal responses. Therefore, using response modality to manipulate arbitrariness of the concept-response mapping always implied a confounding of mapping compatibility with distractor-response compatibility. I tried to meet these objections by using vocal/arbitrary mappings in Experiments 3 and 4, however the results of the last experiment suggest that this was only partly successful. The problem with the vocal/arbitrary mappings was that overall response times were considerably larger, and Stroop effects were weaker than with the manual/arbitrary mapping. In the discussion of the last experiment I speculated that phonological priming might have decayed by the time the mapping rule had been selected. However, it seems desirable to separate the effects of automatic priming and controlled translation.

To more directly dissociate the differential interference effects along the direct route on the one hand and along the semantic and episodic buffer route on the other hand, the current experiment included a manipulation of distractor-response compatibility. To achieve this, arrow and word distractors were used in both response modalities. As has been discussed above, distractor words have fast access to phonological response codes in the vocal response conditions and can lead to a fast, automatic pre-activation of a response set element, which might cause both Stroop interference and facilitation. On the other hand, there is little feature overlap between word distractors and manual-spatial response codes.[49] There is evidence that arrows lead to a similar, 'automatic' preactivation of manual responses as words do for vocal responses (Eimer, 1995, 1997; Wascher, Reinhard, Wauschkuhn, & Verleger, 1999). For example, Eimer (1995) found an early Lateralized Readiness Potential (LRP) activation corresponding to the arrow's

---

[49] This is probably the main reason why the Stroop effect is typically reduced for manual as compared to vocal responses.

direction, which was largely independent of objective cue-response contingencies, and therefore presumably indicates an involuntary, automatic process. Thus, arrow distractors and manual responses seem to share a spatial code, while words are phonologically coded. For distractors that have featural overlap with the response code, priming of the corresponding response set member along the direct, unconditional route is expected to cause relatively large Stroop effects. On the other hand, Stroop effects caused by distractors whose code does not overlap with the response code are limited to the semantic, conditional route, because the distractor code has to be translated before it reaches the response. For example, even a direction word needs to be translated from a graphemical code to activate its spatial semantic before it can affect manual responding. Similarly, the spatial code of an arrow needs to be translated into a phonological code before it can affect vocal responding. Consequently, an expected effect of the manipulation of distractor type is an interaction of distractor type, response modality, and Stroop condition. An incongruent word presumably causes larger interference effects in the vocal than in the manual modality, while the reverse pattern is expected for incongruent arrows.

Using both spatially and phonologically coded distractors has the additional advantage that it allows for a comparison of age effects between the two types of distractors. This is important, because it has been suggested that age effects are far less severe in the lexical than in the nonlexical domain (e.g., Jenkins et al., 2000; Hale & Myerson, 1996). If this is true, then old adults might possibly be more susceptible to interference from verbal than from spatial distractors, because the former are more effectively processed. On the other hand, age differences are expected to be relatively small when the (vocal) response uses a lexical code, and relatively large when the (manual) response uses a spatial code.

The complete design of the current experiment implements the orthogonal manipulation of the within-subjects factors of Mapping (compatible vs. arbitrary), Response modality (vocal vs. manual), Distractor type (word vs. arrow), and Stroop condition (congruent, neutral, incongruent).

Two experiments were run which had an identical design at the level just described. In both experiments, arrows and words were used distractors in both mapping conditions, and also as targets in the compatible mapping condition. However, the experiments differed with regard to the configuration of the stimulus ensembles, and with regard to the target category used in the arbitrary mapping conditions. In Experiment 5(a), target and distractor were spatially separated, as the target was presented above or below the distractor. In Experiment 5(b), target and distractor were integrated on the same object. Experiment 5(a) used symmetric shapes as targets in the arbitrary mapping conditions, while in Experiment 5(b) participants learned to associate colors with directions.

Experiment 5(b) was run because of some difficulties with the stimulus material in Experiment 5(a). However, it turned out that the material used in Experiment 5(b) had its own difficulties. Therefore, the results of the individual experiments will be presented in relatively brief format, and more weight will be given to the subsequent presentation of the results from a combined analysis, including data from both Experiments 5(a) and (b). The combined analysis was performed with the rationale that consistent effects which were due to the experimental design would be highlighted by the increase in power, while unintended side effects due to stimulus peculiarities would vanish in the noise.[50] Figure 13 shows examples of incongruent stimuli in Experiments 5(a) and (b) classified by target-response conceptual compatibility (Mapping) and distractor type.

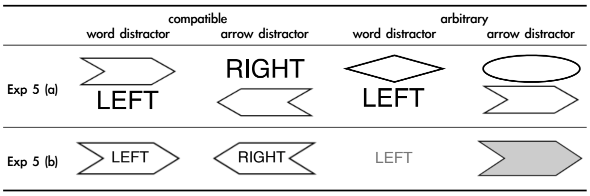


Figure 13. Examples of incongruent stimulus ensembles used in Experiment 5. The upper row shows stimuli from Experiment 5(a), and the lower row stimuli from Experiment 5(b). Columns from left to right depict combinations of (1) arrow target and word distractor, (2) word target and arrow distractor, (3) shape or color target and word distractor, and (4) shape or color target and arrow distractor. Depending on response modality, distractor-response compatibility (DRC) of one and the same ensemble is either high or low. For vocal responding, DRC is high with ensembles (1) and (3). For manual responding, DRC is high with ensembles (2) and (4).

Experiment 5(a): target and distractor spatially separated

Methods

Participants. Twenty-four young and 24 old subjects (*mean age*: 20.2 vs. 69.7, *range*: 18-25 vs. 63-75 years) from the University of Potsdam subject pool participated in Experiment 5(a). Young subjects received DM 10,- and old subjects received DM 15,- per session in an experiment lasting for two sessions of approximately one hour. Unfortunately, due to data collection errors, descriptive data were missing for seven young participants. However, the results for the remaining participants do not at all deviate from the typical pattern. Age groups were comparable with respect to total years of formal education (young, $M = 12.9$, $SD = 2.0$; old, $M = 12.9$, $SD = 4.1$), $t(39) < 1$. Young adults scored much better than old adults in the digit-symbol substitution test (young, $M = 63.5$, $SD = 9.6$; old, $M = 47.2$, $SD = 6.8$), $t(39) = 6.40$, $p < .001$. There was a tendency for old adults to perform better in the MWT-A vocabulary test (young, $M = 31.4$, $SD = 2.4$; old,

---

[50] I would like to thank Dirk Vorberg for pointing me to the justifiability and usefulness of data pooling in an analogue case.

$M = 32.7$, $SD = 2.1$), $t(39) = 1.93$, $p = .061$. All participants were healthy according to self-rating and had normal or corrected-to normal vision.

Design. Unconfounding of response modality and arbitrariness of conceptual S-R mapping was achieved by completely crossing the two. The Response modality factor compared manual and vocal response. The Mapping factor compared compatible and arbitrary mappings of concepts to responses. For the compatible mappings, to obtain conceptually compatible relations with both response modes, stimuli with a pre-experimentally established spatial semantic were used, namely arrows and direction words. For the arbitrary mappings, the association of the target stimulus value and the response was acquired during the course of the experiment.

Furthermore, the previous confounding between concept-response compatibility and distractor-response compatibility was resolved by using direction word as well as arrow distractors in both response modalities. Distractors thus always had a spatial semantic, i.e. they were compatible with the response at the conceptual level. However, depending on response modality they either required translation from a verbal into a spatial code or vice versa, or they could directly activate a response set member due to overlap of the internal code. The manipulation of Distractor type was orthogonal to the Mapping manipulation, which varied the conceptual compatibility of target and response.

Finally, Stroop condition, i.e., congruency of target and distractor semantics, was varied in three steps, like in the experiments reported above. In the congruent condition, target and distractor indicated the same direction, and in the incongruent condition they indicated the opposite direction. In the neutral condition, the distractor was ambivalent with respect to a direction.

To summarize, two Age groups were compared, and for each age group the factorial $2 \times 2 \times 2 \times 3$ within-subjects design orthogonally varied Response modality (manual, vocal), Mapping or conceptual compatibility (compatible, arbitrary), Distractor type (arrow, word), and Stroop condition (congruent, neutral, incongruent).[51]

Stimuli and Procedure. All factors (except Age) were varied within subjects in an experiment lasting for two sessions, each involving 288 warm-up and 864 experimental trials. Response modality was varied between sessions in counterbalanced order, which was matched between age groups. Sensitivity of the voice key was individually calibrated at the beginning of a vocal response session. A session started with four warm-up blocks of 72 target-only trials each to familiarize subjects with the experimental apparatus and to establish the arbitrary mapping. Two of these blocks used arbitrary stimuli, while one of the remaining two warm-up blocks used words and the other used arrows. Thus 144 of 288 warm-up trials were used to build up an association of symmetric geometric shapes (oval and diamond) with left and right responses. Warm-up blocks were presented in the same target order as experimental blocks.

After the warm-up blocks, four blocks of 216 trials each were run, corresponding to each of the four combinations of Mapping and Distractor type. For each level of Mapping, there were two blocks of trials differing in Distractor type, arrow or word. In the compatible conditions, word and arrow targets were paired with distractors from the other category, i.e., arrow targets were paired with word distractors and vice versa. In the arbitrary mapping conditions, shape targets were paired with direction word distractors in one block and with arrow distractors in another block. Possible target-distractor combinations thus included arrow-direction word, direction word-arrow, shape-direction word, and shape-arrow.[52] Each of these combinations was tested with both response modalities.

At the beginning of a block, the current target dimension was specified and a reminder of the mapping rules was presented until the participant indicated that she was ready. Each block was preceded by twelve warmup trials that were discarded from the analysis. Four different orders of combinations of Mapping and Distractor type were used, which were balanced according to a latin square with the constraint that Mapping was alternated between blocks. Half of the subjects started with a compatible Mapping, while the other half started with an arbitrary Mapping. In each block a target was specified by its category membership (e.g., arrow target in block 1, shape target in block 2, word target in block 3, and shape target in block 4). Within each block, Stroop condition was randomized such that compatible, neutral, and incompatible target-distractor pairs were presented on one third of the trials per block. Within each block, all combinations of target direction and Stroop condition appeared equally often (N=36).

A trial started with the presentation of an empty fixation frame consisting of the four corners (extension of 5 pixels, e.g. right and down for the upper left corner) of a virtual rectangle (3.7 by 2.6 °VA), which was presented in white on a black background. Target and distractor simultaneously appeared 900 ms later inside the frame and remained visible until the response was given. After the response, the next trial was immediately initiated by blanking the interior of the fixation frame. An opportunity to pause was given after every 72th trial.

Depending on the type of distractor in the current block, either the words LEFT and RIGHT or leftwards

---

[51] Note that the Distractor type factor can also be rephrased as a distractor-response code compatibility factor by relabelling its levels depending on response modality. For word distractors, code compatibility is high with vocal and low with manual responses, and the reverse applies for arrow distractors.

Note also that arrows and words were used as both targets and distractors in the compatible conditions, while they were only distractors in the arbitrary conditions. Thus in the compatible conditions, whenever distractor-response compatibility was low, *target*-response compatibility was particularly high, because the target could activate the response along the direct route.

[52] Note that target and distractor switched their role only between blocks using a compatible concept-response mapping.

and rightwards pointing arrows were used as distractors in the congruent and incongruent Stroop conditions. The neutral Stroop conditions were generated by using either double-sided arrow distractors or the word MITTE (German for 'center'), depending on whether the type of distractor in the current block was an arrow or a word, respectively.

Target-distractor ensembles subtended a visual angle of 3.3 by 2.2 °VA (horizontally by vertically) when viewed from a distance of 70 cm. The bounding box of arrows and shapes had an extension of 3.1 by 1.0 °VA. Words were printed in capital letters using the Geneva font at a font size of 36 points, which was chosen such that the horizontal extension of the longer target word (RECHTS) approximately matched the horizontal extension of the arrows.

To discourage selection based on spatial location, the target randomly appeared above or below the distractor. The whole target-distractor ensemble was centered on fixation.

Manual responses were registered using the response keys on the CMU button box, whose colors had been hidden by sticking grey tape on them. Vocal responses were registered by the voice key in the CMU button box.

Results

Because the design was rather complex, to facilitate presentation and perception of the results, I will focus on the theoretically relevant interactions. Specifically, the focus of interest is whether age differences in the Stroop effect are modulated by Mapping, Response modality, or a combination of both. Data were analyzed using a $2 \times 2 \times 2 \times 2 \times 3$ ANOVA with Age as between subjects factor, and Response modality, Mapping, Distractor type, and Stroop condition as within subjects factors.

Reaction times. Mean reaction times by Age group, Mapping, Response modality, Distractor type, and Stroop condition after removal of outliers (2.1%) and errors (3.06%) are presented in Table 5. Neither of the critical interactions reached significance, Age by Response modality by Stroop condition, $F(2,92) = 1.28$, $p = .282$, $MS_e = 173.64$, Age by Mapping by Stroop condition, $F(2,92) = 2.49$, $p = .088$, $MS_e = 174.57$, Age by Response modality by Mapping by Stroop condition, $F(2,92) = 1.70$, $p = .189$, $MS_e = 179.73$. However, detailed analysis of the Stroop contrasts reveals one significant effect, namely, there was a triple interaction of Mapping, Age, and the Stroop interference contrast, $F(1,46) = 4.62$, $p = .037$, $MS_e = 363.23$. This effect is in the previously observed direction: the age difference in interference is larger with arbitrary than with compatible mappings. The interaction was not qualified by any higher order interactions.

In particular, the modulation of the age difference in Stroop interference by Mapping was not affected by the degree of overlap of distractor and response, $F < 1$, although strong age-equivalent effects of stimulus-response compatibility due to code overlap were observed. The fact that overlap of stimulus and response codes leads to fast responses is indicated by the strong interaction of Response modality and Distractor type, $F(1,46) = 66.16$, $p < .001$, $MS_e = 1692.02$, which is entirely limited to the compatible mapping (for problems with the interpretation of Distractor type effects with the arbitrary mapping, see below). Recall that in the compatible mapping conditions, a change in distractor type also meant a change in target type. Thus with a vocal response and arrow distractors, the 'reverse spatial Stroop task' consisted in reading a word while ignoring a simultaneously presented arrow. Similarly, responding with a keypress to an arrow target accompanied by word distractors can also be considered a reverse spatial Stroop task. For the compatible mappings, the interaction indicates that translation between internal codes comes with a cost: responding is fast when a phonological code can be used for vocal responses, or a spatial code for manual responses, and it is slower in the opposite pairings. The size of this interaction was equivalent between age groups, $F < 1$, which might indicate that the speed of the automatic activation of a response code is not affected by an age-related decrement. Stimulus-response compatibility also affects the Stroop effect: there was a strong interaction of Response modality, Distractor type, and Stroop condition, $F(2,92) = 43.41$, $p < .001$, $MS_e = 235.20$, which indicates that arrow distractors cause more interference than word distractors when responding manually, while word distractors cause more interference (and facilitation) than arrow distractors when responding vocally. Thus, the degree of overlap between distractor and response codes does affect the size of the Stroop effect, which supports a dual-route conception. This interaction was not modulated by age, $F < 1$, again supporting the view that the locus of age effects is at a more central, cognitive stage. Averaged across response modalities, the interaction of Distractor type and Stroop condition was also significant, $F(2,92) = 12.33$, $p < .001$, $MS_e = 190.44$. This was caused by higher facilitation with word as compared to arrow distractors, $F(1,46) = 13.85$, $p = .001$, $MS_e = 490.22$, while interference was the same for both distractor types, $F < 1$, and might indicate that reading is more automatic than manually responding to an arrow.

However, interpretation of all interactions involving Distractor type is complicated by results showing that there was a problem with the stimulus material, as was mentioned in the introduction to the current experiment. The very strong interaction effect of Mapping and Distractor type, $F(1,46) = 199.72$, $p < .001$, $MS_e = 2134.91$, indicates that the particular choice of stimulus material was problematic. Responding on trials with shape targets and arrow distractors was 49 ms slower than on trials with shape targets and word distractors, and 72 ms slower than on trials with word targets and arrow distractors. In fact, the main effect of Mapping, $F(1,46) = 113.64$, $p < .001$, $MS_e = 2803.58$, meaning that responses with compatible mappings were 33 ms faster than with arbitrary mappings, might be entirely due to the interaction, because for trials with word

Table 5

Means and standard errors [ms] for reaction time (columns 1-8), and mean error percentages (columns 9-12) in Experiment 5(a), broken up by Mapping, Age, Response modality, Distractor type, and Stroop condition.

			mean reaction time (s.e.) [ms]				*percent errors*			
			compatible		arbitrary		compatible		arbitrary	
Response	*Distractor*	*Stroop*	young	old	young	old	young	old	young	old
vocal	word	congr.	459 (9.3)	511 (11.6)	435 (11.2)	499 (12.6)	1.4	1.3	1.5	1.7
		neutr.	473 (9.3)	532 (15.7)	450 (14.5)	520 (14.7)	2.4	2.0	1.7	2.4
		incongr.	487 (13.0)	542 (15.8)	459 (14.1)	528 (15.6)	5.6	4.2	4.5	5.3
	arrow	congr.	413 (8.0)	455 (10.5)	492 (13.6)	563 (14.4)	1.8	0.5	2.3	2.7
		neutr.	415 (7.8)	454 (10.6)	492 (12.7)	562 (15.0)	2.3	0.8	2.5	2.9
		incongr.	416 (7.9)	456 (10.7)	499 (12.9)	573 (16.0)	2.3	0.9	3.2	3.3
manual	word	congr.	380 (7.2)	460 (8.9)	378 (8.2)	466 (9.9)	4.0	2.0	3.2	1.0
		neutr.	382 (7.6)	464 (10.4)	385 (9.1)	475 (10.4)	4.2	2.1	4.5	1.8
		incongr.	381 (6.9)	471 (10.6)	392 (9.5)	491 (12.4)	3.9	2.4	5.3	3.1
	arrow	congr.	389 (7.7)	463 (6.6)	422 (7.8)	523 (9.6)	4.3	2.0	4.0	2.7
		neutr.	395 (8.2)	472 (7.2)	429 (9.3)	524 (9.8)	5.4	2.6	6.4	2.5
		incongr.	408 (9.3)	476 (9.1)	435 (9.8)	549 (11.0)	6.0	3.7	5.7	3.8

distractors, responses in the arbitrary mapping condition were actually *faster* than with the compatible mapping.[53] Although the interaction was further modulated by the significant triple interaction of Mapping, Distractor type, and Response modality $F(1,46) = 65.96$, $p < .001$, $MS_e = 1524.30$, this merely means that the unintended slowing of vocal responses on trials with shape/arrow stimuli was even larger than the still very sizeable effect in the manual modality. Note that the significant interaction of Age, Mapping, and Stroop interference that was discussed above is not directly affected by the strange behavior of the shape/arrow stimuli, because it was not modulated by higher order interactions, in particular not by and interaction involving Distractor type. The prominent position of the shape/arrow ensembles is very likely due to the fact that this was the only stimulus ensemble without a large difference in spatial frequency between target and distractor. The other three combinations all included a pairing of a word, which is a stimulus with a relatively high spatial frequency, and an arrow or shape, both of which are relatively low spatial frequency stimuli. Thus in all but the problematic shape/arrow condition, the attentional filter could be tuned to select the target location based on spatial frequency information alone. This bears resemblance to a visual search task with feature pop-out, where selection is accomplished without a capacity limit (which is sometimes called preattentional). In contrast, selecting the target in the shape/arrow ensemble probably requires a more controlled search, which takes some time even in the very limited case of only one distractor.

Having discussed the problematic aspects of the current task, for the sake of completeness, I will report the additional significant effects in the following paragraph. All main effects were significant, Age, $F(1,46) = 33.87$, $p < .001$, $MS_e = 15503.96$, Response modality, $F(1,46) = 59.37$, $p < .001$, $MS_e = 9709.44$, Distractor type, $F(1,46) = 15.64$, $p < .001$, $MS_e = 2015.30$, and Stroop condition, $F(2,92) = 78.58$, $p < .001$, $MS_e = 301.30$ with both facilitation, $F(1,46) = 46.70$, $p < .001$, $MS_e = 474.90$ and interference, $F(1,46) = 65.99$, $p < .001$, $MS_e = 3820.01$ (for Mapping, see above). Old adults were 74 ms slower than young adults, manual responses were 45 ms faster than vocal responses, and fairly small, but reliable Stroop facilitation and interference effects of 8 ms each were obtained. The main effect of distractor type means that responding was about 11 ms slower on trials with arrow distractors than on trials with word distractors.

Many of the findings from the previous experiments were replicated, with one notable exception: there was no interaction of Age and Stroop condition, $F(2,92) = 1.90$, $p = .155$, $MS_e = 301.30$, i.e. averaged across conditions young and old adults did not show any difference in the Stroop effect.

Like in the previous experiments, Age interacted with Response modality, $F(1,46) = 6.80$, $p = .012$, $MS_e = 9709.44$. Both age groups were faster when responding manually, but young adults improved more. However, compared to the previous experiments, the size of the interaction effect was rather small (32 ms). Age also interacted with Mapping, $F(1,46) = 11.24$, $p = .002$, $MS_e = 2803.58$: the Mapping effect was 23 ms for young adults, and 46 ms for old adults.

Mapping and Stroop condition also interacted, $F(2,92) = 4.39$, $p = .015$, $MS_e = 174.57$, facilitation n.s. ($F < 1$), interference, $F(1,46) = 7.40$, $p = .009$, $MS_e = 363.23$: Stroop effects were somewhat larger with the arbitrary mapping.

Finally, Response modality interacted with Mapping, $F(1,46) = 6.09$, $p = .017$, $MS_e = 1347.84$, which constitutes a difficult to interpret interaction. The Mapping effect was somewhat larger with vocal than with manual

---

[53] This might be explained by the fact that target and distractor switched their roles in perceptually identical stimulus ensembles only between compatible mapping conditions.

responses. However, this occurred only with arrow distractors, while the mapping effect was actually reversed for vocal, but not manual, responses for ensembles with word distractors. Thus this interaction appears to be fully qualified by the problematic triple interaction with distractor type mentioned above.

Errors. Because of the low overall error rate (3.1 %), results will only briefly be described. None of the theoretically most interesting triple interactions involving Age, Stroop condition, and Mapping or Response modality was significant in the analysis of error rates, and neither was the interaction involving all four mentioned factors. In fact, in addition to the Age main effect, $F(1,46) = 12.86$, $p = .001$, $MS_e = 1.23\text{e-}3$, which indicates that old adults made less errors than young adults ($Ms = 2.6$ *vs*. 3.7%, respectively), there were only two significant interactions involving Age. Age interacted with Response modality, $F(1,46) = 9.67$, $p = .003$, $MS_e = 2.8\text{e-}3$, because young adults made significantly more errors in the manual than in the vocal modality ($Ms = 4.7$ *vs*. 2.6%, respectively), while for old adults error rates were equivalent in both response modalities. The error and reaction time effects for response modality and age are in the opposite direction. A speed-accuracy tradeoff explanation is thus possible: It could be that young adults chose a more liberal response criterion when responding manually.

The second interaction involving Age was a triple interaction of Age, Response modality, and Mapping, $F(1,46) = 5.69$, $p = .021$, $MS_e = 8.75\text{e-}4$. For young adults, the error rate was modulated by Response modality, but not by Mapping. For old adults, the error rate was not modulated by Response modality on an overall level, but Mapping had an influence only in the vocal modality, where twice as many errors were produced with the arbitrary than with the compatible mapping.

There was a cluster of interactions involving the factor Distractor type. Distractor type interacted with Stroop condition, $F(2,92) = 9.84$, $p < .001$, $MS_e = 4.73\text{e-}4$, and with Response modality, $F(1,46) = 15.96$, $p < .001$, $MS_e = 1.27\text{e-}3$, and there was a triple interaction involving Distractor type, Response modality, and Stroop condition, $F(2,92) = 14.83$, $p < .001$, $MS_e = 4.47\text{e-}4$. Overall, Stroop interference effects were larger with word than with arrow distractors. Arrow distractors caused more errors than word distractors with manual responding, and less errors than word distractors with vocal responding. The latter fact was largely due to the rather high error rate elicited by incongruent word distractors in the vocal response condition. For neutral and congruent distractors in the vocal response modality, there was no difference in error rates. The triple interaction also indicates that the large interference effects caused by word distractors are limited to vocal responding: in the manual modality, interference and facilitation effects were equivalent between distractor types. However, arrow distractors caused a larger Stroop effect in the manual than in the vocal modality, while the reverse pattern was obtained for word distractors. Taken together, the degree of overlap of distractor and response code apparently modulated the speed of responding and the size of the Stroop effect.

Finally, the four-way interaction of Mapping, Response modality, Distractor type, and Stroop condition was significant, $F(2,92) = 3.32$, $p = .041$, $MS_e = 3.55\text{e-}4$. Because this might have partly been caused by the problematic shape/arrow ensemble, I will only describe the pattern of the interaction for ensembles with word distractors.[54] The size of the Stroop effect caused by word distractors was large and equivalent between Mappings in the vocal modality, and it was also large in the manual modality with the arbitrary Mapping, but it was negligible in the manual modality with the compatible mapping. This again appears to support the hypothesis that arrow targets have a direct access to a manual response code.

Discussion. I will postpone discussion of the results until the results of Experiment 5(b) and the combined analyses (including analyses of proportional scores) have been presented. The reader should keep in mind that most of the 'interesting' effects in Experiment 5(a) were not significant, with the exception of the interaction of Age, Mapping, and Stroop interference. However, the size of the interaction was fairly small. Because of the problems with the stimulus material, a second experiment was planned using an identical design, but different target-distractor ensembles.

Experiment 5(b): target and distractor spatially integrated

Participants. Sixteen young (age $M = 20.1$ years, $range = 15 - 25$) and 16 old adults ($M = 70.2$, $range = 67 - 73$) participated in Experiment 5(b). Age groups did not differ in total years of formal education (young, $M = 11.7$, $SD = 3.9$; old, $M = 12.5$, $SD = 3.0$), $t(30) < 1$. Unfortunately, due to data collection errors, MWT-A data were missing for one old and three young participants, and DSS performance data were missing for two young participants. However, the results for the remaining participants do not at all deviate from the typical pattern. Young adults achieved a much higher score than old adults in the digit-symbol substitution test (young, $M = 64.6$, $SD = 12.7$; old $M = 47.6$, $SD = 9.3$ points), $t(28) = 4.2$, $p < .001$. Old adults performed slightly, but significantly better on the MWT-A vocabulary test (young, $M = 30.1$, $SD = 2.4$; old $M = 32.1$, $SD = 1.2$ points), $t(17) = 2.76$, $p = .013$.[55] All participants were healthy according to self-rating and had normal or corrected-to normal vision.

Stimuli and Procedure. In Experiment 5(b) the same logic as in Experiment 5(a) was applied, but targets in the arbitrary mapping conditions were changed to colors, so that arrow or word distractors were presented in

---

[54] This seems justifiable in particular because if only trials with word distractors are analyzed, a significant three-way interaction of Mapping, Response modality, and Stroop condition is obtained, $F(2,92) = 3.39$, $p = .038$, $MS_e = 4.07\text{e-}4$.

[55] The t-value and the degrees of freedom were adjusted because of unequal variance.

the target color. Because target and distractor in the arbitrary condition were thus presented on the same object, the stimuli in the compatible mapping conditions were also changed. In an attempt to integrate words and arrows on the same object, words were presented inside a surrounding arrow. Targets were presented centrally at fixation. In the compatible mapping conditions, the word was presented in the black background color inside a white arrow. Identical stimulus ensembles were used in blocks with word targets and arrow distractors, and with arrow targets and word distractors. In the arbitrary mapping conditions, depending on distractor type, either the word or the arrow was presented in cyan or magenta color. Again, the responses consisted in pronouncing 'left' or 'right', or in pressing a left or right key. Arrows subtended a visual angle of 4.5 x 1.3 °VA (horizontally by vertically), and words subtended a visual angle of 2.6 x 0.8 °VA when viewed from a distance of 70 cm. The empty fixation frame at the beginning of a trial had an extension of 5.8 by 1.9 °VA. All other aspects were unchanged from Experiment 5(a).

Results

Reaction times. After removal of outliers (2.1%) and errors (5.8%), aggregated mean reaction times (see table 6) were subjected to a 2 (Age) × 2 (Response modality) × 2 (Mapping) × 2 (Distractor type) × 2 (Stroop condition) repeated measures analysis of variance with Age as between subjects factor.

With the current stimuli and design, there were sizeable Stroop effects $F(2,60) = 70.46$, $p < .001$, $MS_e = 394.96$, both in interference and facilitation. Although on an overall level, there was a tendency for an increased Stroop effect for old adults, as indicated by the marginal Age×Stroop condition interaction, $F(2,60) = 3.15$, $p = .050$, $MS_e = 394.96$, the critical interaction of Age, Mapping, and Stroop condition did not reach significance, $F(2,60) = 1.50$, $p = .231$, $MS_e = 261.41$. It should be noted, though, that numerically the age difference in the Stroop effect was 3 ms with the compatible (Stroop effect 16 vs. 19 ms, young vs. old) and 12 ms with the arbitrary mapping (18 vs. 30 ms), hence, in the previously observed direction.

With regard to the question whether age differences in the Stroop effect are different between response modalities, there was a significant main effect of Response Modality, $F(1,30) = 122.91$, $p < .001$, $MS_e = 6019.97$, and like in the previous experiments, Age and Response modality interacted, $F(1,30) = 4.72$, $p = .038$, $MS_e = 6019.97$. Furthermore, Response modality interacted with Stroop condition, $F(2,60) = 5.50$, $p = .006$, $MS_e = 250.41$. However, the critical interaction of Age, Response modality, and Stroop condition was far from significant, $F < 1$. Both age groups were faster with manual than with vocal responses, and for young adults the manual advantage was particularly large (response modality effect of 74 vs. 50 ms, young vs. old). Aggregated across levels of Distractor type and Mapping, Stroop effects were larger in the vocal than in the manual modality (Ms $= 25$ vs. 16 ms, respectively). The age difference in the Stroop effect was 8 ms with vocal responses, and 7 ms with manual responses. Thus there was absolutely no sign of enhanced Age×Stroop effects in the manual modality.

In fact, in the current experiment only a single factor modulated the age difference in the Stroop effect, namely the type of distractor, as indicated by the significant three-way interaction of Age, Distractor type, and Stroop condition, $F(2,60) = 7.74$, $p = .001$, $MS_e = 191.44$ in the absence of interactions of Age and Distractor type, or of Distractor type and Stroop condition. Old adults were more affected by word distractors than young adults (mean Stroop effect 11 vs. 27 ms, young vs. old), while there was no age difference in the Stroop effect caused by arrow distractors (24 vs. 22 ms).

Like in Experiment 5(a), strong crossover interactions of Response modality and Distractor type, $F(1,30) = 83.37$, $p < .001$, $MS_e = 1281.34$, and of Response modality, Distractor type, and Stroop condition, $F(2,60) = 53.21$, $p < .001$, $MS_e = 160.22$, were obtained, which indicate that direct-route stimulus-response compatibility is a major determinant of Stroop interference and facilitation. Manual responding was faster with word distractors, and vocal responding was faster with arrow distractors. The Stroop effect was large with word distractors and vocal responding and with arrow distractors and manual responding (35 and 30 ms, respectively), while the 'reverse spatial Stroop effect' was small with arrow distractors and vocal responding, and absent with word distractors and manual responding (16 and 3 ms).

However, as with Experiment 5(a), there was a problem with the choice of stimulus material, albeit a different one. This problem becomes obvious through the interactions of Mapping × Distractor type, $F(1,30) = 69.79$, $p < .001$, $MS_e = 1670.13$, and of Response modality × Mapping × Distractor type, $F(1,30) = 49.77$, $p < .001$, $MS_e = 907.99$. This latter interaction qualified both the main effect of Mapping, $F(1,30) = 19.63$, $p < .001$, $MS_e = 2554.00$ and the interaction of Response modality × Mapping, $F(1,30) = 75.07$, $p < .001$, $MS_e = 828.54$.

The interaction of Mapping and Distractor type means that a (small, 7.5 ms) Mapping effect in the expected direction was observed with word distractors, but a (large, 41 ms) reverse Mapping effect was observed with arrow distractors, for which responding with the arbitrary mappings was faster than with the compatible mappings. The direction of this interaction effect was consistent between Response modalities, however, the size was much larger in the manual (6 vs. -74 ms, Mapping effect word vs. arrow distractors) than in the vocal modality (11 vs. -7 ms), as indicated by the triple interaction of Mapping, Response modality, and Distractor type. Although it is not entirely clear what is responsible for this pattern, I think that it is likely caused by a combination of two effects related to the stimulus material.

First, responding to the word/arrow ensemble used in the compatible conditions was relatively slow when the

Table 6
Means and standard errors [ms] for reaction time (columns 1-8), and mean error percentages (columns 9-12) in Experiment 5(b), broken up by Age, Mapping, Response modality, Distractor type, and Stroop condition.

			mean reaction time (s.e.) [ms]				*percent errors*			
			compatible		arbitrary		compatible		arbitrary	
Response	*Distractor*	*Stroop*	young	old	young	old	young	old	young	old
vocal	word	congr.	438 (8.4)	500 (12.6)	448 (7.5)	495 (17.2)	1.0	1.2	5.1	3.1
		neutr.	446 (12.0)	513 (14.2)	465 (12.4)	516 (18.7)	1.1	1.4	5.2	3.0
		incongr.	456 (12.3)	533 (13.5)	479 (10.2)	553 (24.8)	4.4	3.5	6.5	6.4
	arrow	congr.	432 (7.9)	488 (12.6)	425 (10.8)	476 (16.8)	1.7	1.1	3.0	3.6
		neutr.	443 (9.2)	494 (12.6)	435 (10.1)	481 (15.9)	2.4	1.0	2.3	4.0
		incongr.	446 (9.9)	497 (12.7)	449 (12.6)	494 (17.6)	2.2	0.9	3.8	5.3
manual	word	congr.	354 (6.8)	439 (8.9)	364 (7.5)	433 (10.4)	4.2	1.4	3.1	1.7
		neutr.	354 (7.1)	442 (8.2)	367 (9.7)	447 (15.2)	4.5	1.6	3.7	1.2
		incongr.	354 (7.2)	442 (8.4)	361 (9.1)	451 (17.4)	4.3	2.2	4.3	3.6
	arrow	congr.	406 (10.3)	488 (10.0)	336 (7.1)	412 (9.7)	2.4	1.7	2.3	0.8
		neutr.	431 (12.7)	499 (10.9)	351 (7.3)	425 (8.9)	5.8	2.4	3.5	1.9
		incongr.	442 (12.5)	513 (13.3)	362 (8.7)	445 (13.3)	9.7	4.0	7.1	2.8

word was the target. This problem might be related to figure/ground perception: Because the word was printed in the background color, it might have appeared as part of the background, hence reactions to word targets in the compatible condition might have required some extra time to interpret the word as the figure, and the arrow as background. This conjecture is supported both by the size of the distractor type effect for the compatible mappings in the manual response modality (462 vs. 397 ms, word/arrow vs. arrow/word). If only manual responses are compared, compatible *word/arrow* ensembles constituted an extreme outlier, leading to a reaction time that was 65 ms above the mean of the other three (arrow/word, color/word, and color/arrow) ensembles, whereas the maximum reaction time difference in pairwise comparisons between the latter was 10 ms.[56]

Second, the arbitrary color/arrow ensemble had a somewhat special status in the current experiment, because it was the only ensemble that did not have high spatial frequency components at the center of the screen. The target feature color was carried by a rather large arrow, thus relatively high color 'energy' was available with colored arrows. This might have led to particularly fast responses to that ensemble, thereby contributing to the reverse Mapping effect, which was limited to the comparison of word/arrow and color/arrow ensembles. Indeed, for arbitrary mappings, responding with the color/arrow ensemble was generally faster than with the color/word ensemble.

Errors. Errors were analyzed using the same 2 (Age) × 2 (Response modality) × 2 (Mapping) × 2 (Distractor type) × 3 (Stroop condition) ANOVA as reaction times. There were two significant main effects, Age, $F(1,30) = 5.76$, $p = .023$, $MS_e = 2.26\text{e-}3$, and Stroop condition, $F(2,60) = 44.19$, $p < .001$, $MS_e = 6.98\text{e-}4$. Young adults made more errors than old adults, and Stroop interference and facilitation were obtained.

The critical interaction of Age, Mapping, and Stroop condition that failed to reach significance in the reaction time analysis was significant in the error analysis, $F(2,60) = 3.20$, $p = .048$, $MS_e = 3.68\text{e-}4$. For young adults, the Stroop effect in errors was larger with compatible than with arbitrary Mappings, while for old adults, it was numerically larger with arbitrary Mappings. Furthermore, with the compatible Mappings, young adults' Stroop effect was larger than old adults', while the reverse was true with arbitrary Mappings. This pattern hints at some specific age-related deficit with arbitrary rules, that leads to an age-relative amplification of Stroop effects. Although the size of the Stroop effect for old adults was not larger with the arbitrary than with the compatible Mappings, the age difference increased, because young adults produced a significantly smaller Stroop effect with arbitrary than with compatible Mappings.

Like in reaction times, there was a strong interaction of Age, Distractor type, and Stroop condition, $F(2,60) = 7.03$, $p = .002$, $MS_e = 4.10\text{e-}4$. Old adults experienced more interference from word distractors than young adults, who experienced more interference from arrow distractors than old adults. The interaction was crossover, as young adults produced a larger Stroop effect than old adults with arrow distractors, and old

---

[56] Further observations supporting the conjecture that the compatible word/arrow ensemble had a somewhat special status comes from two observations: First, there was no reading benefit for word/arrow ensembles in comparison to arrow/word ensembles in the vocal response modality: vocal responding with word/arrow ensembles (467 ms) was not much faster than vocal responding with arrow/word ensembles (481 ms), although the former simply required reading of the word. Second, the results from the previous experiments lead to the expectation of a manual response benefit at least for young adults. In the present experiment, the size of this benefit was only very weak in the compatible *word/arrow* conditions, and much stronger in the compatible word arrow/word conditions.

adults produced a larger Stroop effect than young adults with word distractors. Again, the reasons could be related to age-differential pre-experimental training with arrow-to-key associations, however, this interpretation remains speculative because results might equally well suggest that young adults prefer a spatially mediated rehearsal strategy, while old adults prefer a verbal strategy.[57]

A further result that mirrored results from the reaction time analysis was the interaction of Response modality, Distractor type, and Stroop condition, $F(2,60) = 18.72$, $p < .001$, $MS_e = 5.61\text{e-}4$. Stroop effects were larger with distractors that overlapped with the response codes than with distractors that needed to be translated. This was a crossover interaction with respect to the Stroop effect. The ratio of incongruent to congruent condition error rates for vocal responding was 2.0 with word distractors and 1.3 for with arrow distractors, while for manual responding the ratio was 1.4 with word and 3.2 with arrow distractors. An additional contribution to the interaction stems from the fact that large facilitation effects were only observed with manual responses and arrow distractors. This interaction was further modulated by Mapping, as indicated by the significant interaction of Response modality, Mapping, Distractor type, and Stroop condition, $F(2,60) = 4.82$, $p = .012$, $MS_e = 4.42\text{e-}4$. In the compatible mapping conditions, Stroop effects in error rate were negligible when the target was response-compatible (i.e. with low-overlap distractors, vocal: word/arrow and manual: arrow/word), while they were substantial when the target had to be translated (and the distractor was response-compatible: vocal: arrow/word, manual: word/arrow). Stroop effects in the latter conditions were larger than the sizeable Stroop effects in the arbitrary Mapping conditions, where the triple interaction of Response modality, Distractor type, and Stroop condition was nevertheless observed.

Unlike in reaction times, the main effect of Mapping, $F(1,30) = 4.36$, $p = .045$, $MS_e = 3.37\text{e-}3$ does not indicate a reversal of the Mapping effect, but more errors were observed with arbitrary than with compatible mappings. However, recall that in the reaction time analysis, the reversal was caused by stimulus ensembles featuring arrow distractors, and mainly limited to the manual modality. While in the error analysis, the interactions of Mapping and Distractor type, or of Response modality Mapping and Distractor type failed to reach significance, two effects reflect problems with the particular choice of stimuli. First, part of the four-way interaction referred to in the last paragraph was caused by a reversal of the mapping effect with manual responding and arrow distractors. Second, there was a significant interaction of Response modality and Mapping, $F(1,30) = 11.28$, $p = .002$, $MS_e = 4.16\text{e-}3$. The Mapping effect was in the standard direction only for vocal responses, while it was slightly reversed for manual responses (mainly due to the reversal in the case of arrow distractors).

Both experiments combined

Because the conceptual design of the experiments was identical, in an additional analysis data from both experiments were pooled, i.e. the analysis was repeated with combined data from Experiment 5(a) and (b), including experiment as an additional between subjects factor. The analysis of variance thus included the between subjects factors of Experiment and Age, and the within subjects factors of Mapping, Response modality, Distractor type, and Stroop condition.[58] The motivation for a combined analysis was twofold. First, increasing the number of subjects should lead to an increase in power, which was welcome since all of the Stroop congruency effects were rather small numerically. Second, because the results from the single experiments were somewhat muddled, it was expected that this combined analysis would reveal the consistent effects, while effects due to stimulus peculiarities would vanish in the noise.

The strategy in interpretation of the results was as follows: an effect is considered to be consistent if its direction and strength was the same between experiments, which is indicated by a lack of interaction with the Experiment factor. For effects that did interact with Experiment, a distinction was made between effects that only differed in strength between experiments, but whose direction was the same, and effects that switched direction between experiments, leading to a crossover interaction with Experiment. The former are relevant, although not consistent. The latter can be considered artifacts of stimulus or other differences between experiments, because the conceptual design of the two experiments was identical. Hence, these effects are of questionable relevance with respect to the research question.

Reaction times. Apart from the Age main effect, $F(1,76) = 54.42$, $p < .001$, $MS_e = 14106.62$, only three effects involving Age did not interact with Experiment, namely the interactions of Age and Response modality, $F(1,76) = 9.63$, $p = .003$, $MS_e = 8298.28$, of Age and Stroop condition, $F(2,152) = 4.65$, $p = .011$, $MS_e = 332.55$, and the critical interaction of Age, Mapping, and Stroop condition, $F(2,152) = 3.94$, $p = .021$, $MS_e = 215.21$. Old adults were slower than young adults, and the age difference was larger with manual than with vocal responding (mean age effect of 84 vs. 58 ms, respectively), which was slower than manual responding in both age groups. Old adults produced larger Stroop effects than young adults. However, the Age × Stroop condition interaction is qualified by the

---

[57] Strictly, this latter interpretation would be supported by an interaction of Age, Mapping, Distractor type, and Stroop condition. Although the interaction was not significant, $F(2,60) = 2.05$, $p = .137$, $MS_e = 4.85\text{e-}3$, the age-differential effect of word distractors on Stroop interference is limited to the arbitrary mappings.

[58] The SPSS® GLM procedure that was used to evaluate the effects uses a weighted least squares algorithm to estimate model parameters, so that the difference in sample size between the experiments is taken into account.

interaction of the two factors and Mapping, which—as analyses using repeated contrasts for Stroop condition show—is due to old adults' larger interference effects, $F(1,76) = 4.23$, $p = .043$, $MS_e = 488.81$, while there was no Age difference in facilitation, $F < 1$: The age difference in the Stroop effect was relatively large with arbitrary mappings (mean Stroop effect 17 vs. 27 ms, young vs. old), while there was no age difference in the Stroop effect with compatible mappings (15 vs. 16 ms). Separate ANOVAs for each type of Mapping show that the interaction of Age and Stroop condition is highly significant for the arbitrary mappings, $F(2,152) = 6.70$, $p = .002$, $MS_e = 349.23$, while it is far from significant for the compatible mappings, $F < 1, p = .757$. Separate analyses for each Age group reveal that the Stroop effect is enhanced under arbitrary mapping conditions in the old group, $F(2,76) = 8.66$, $p < .001$, $MS_e = 294.14$, while Mapping does not modulate the Stroop effect in the young group, $F < 1$, $p = .526$. Therefore the interaction of Mapping and Stroop condition, $F(2,152) = 8.30$, $p < .001$, $MS_e = 215.21$, that was observed independent of Experiment seems to be largely due to the effect in the older group. Interestingly, in the current analyses the triple interaction of Age, Mapping, and Stroop condition was obtained in the absence of an Age × Mapping interaction, $F(1,76) = 2.47$, $p = .120$, $MS_e = 2723.50$.[59] Thus with regard to the first question of interest, we can conclude that in the present paradigm(s), an Age effect in Stroop interference only occurs with arbitrary mapping conditions. In both experiments, the age effect in Stroop interference was (numerically) higher with the arbitrary than with the compatible mapping. The upper left panel of Figure 14 shows the pattern of reaction times for the critical interaction, which was not further modulated by Experiment, Response modality, or a combination of both. The set of interactions involving Mapping and Experiment that were independent of Stroop condition and only slightly mediated by Age do not in my opinion disqualify the results concerning the critical interaction of Age, Mapping, and Stroop condition. In fact, one could argue that despite the inconsistent effects of Mapping between experiments, the aspect of the Mapping manipulation that led to the interaction with Age and Stroop condition was consistent between experiments. In summary, old adults are subject to enhanced interference even under the relatively small memory load imposed by holding on-line only two stimulus-response pairings.

The second research question addressed a possible modulation of Age effects in Stroop interference by Response modality. As shown in the lower left panel of Figure 14, the critical three-way interaction of Age, Response modality, and Stroop condition was far from significant, $F(2,152) = .25$, $p = .778$, $MS_e = 191.40$, and was not modulated by Experiment. Although the size of the Stroop effect differed between Response modalities, $F(2,152) = 4.89$, $p = .009$, $MS_e = 191.40$, and the interaction of Response modality and Stroop condition also interacted with Experiment, $F(2,152) = 4.33$, $p = .015$, $MS_e = 191.40$, indicating that the Stroop effect was larger with vocal than with manual responding, at least in Experiment 5(b)[60], this pattern of interactions was equivalent in both Age groups. Thus it appears that response modality per se is not a mediator of Age differences in the Stroop effect.

A third age-related question was raised by the results of the individual experiments, namely whether arrow and word distractors differ in their influence between Age groups, either on the Stroop effect or on maintenance of the arbitrary Mapping. The effect of Distractor type on the maintenance of the arbitrary mapping is difficult to evaluate, because it was inconsistent between experiments, as indicated by the crossover interaction of Mapping, Distractor type, and Experiment, $F(1,76) = 10.23$, $p = .002$, $MS_e = 1954.75$, which modulated the interaction of Mapping and Distractor type, $F(1,76) = 233.45$, $p < .001$, $MS_e = 1954.75$. Mapping effects in Experiment 5(a) were nearly absent with word distractors and large and in the expected direction with arrow distractors. In contrast, in Experiment 5(b), Mapping effects were small, but in the expected direction with word distractors, and large, but in the reverse direction with arrow distractors. Speculations about the possible causes were brought forward in the results section of the individual experiments. Although Age did not interact with either of the two interactions, the interpretation of a possible age equivalence of the influence of Distractor type on the maintenance of arbitrary Mappings is problematic because of the inconsistent pattern of effects between experiments, and will therefore be no further discussed. With regard to a possible age-differential pattern of Stroop effects mediated by distractor type, a substantial interaction of Age, Distractor type, and Stroop condition was observed, $F(2,152) = 9.01$, $p < .001$, $MS_e = 182.03$, which was only marginally modulated by Experiment,

---

[59] The failure to obtain the latter is probably caused by the fact that the interaction was different between experiments, as indicated by the significant triple interaction of Experiment, Age, and Mapping, $F(1,76) = 6.98$, $p = .010$, $MS_e = 2723.50$. In Experiment 5(a), where a significant Mapping effect in the expected direction was obtained, the interaction of Mapping with Age was also significant, whereas in Experiment 5(b), where the main effect of Mapping actually indicated a reversal of the typical pattern, no interaction with Age was observed. On an overall level, the Mapping main effect points in the expected direction, $F(1,76) = 11.96$, $p = .001$, $MS_e = 2723.50$, with arbitrary Mappings causing slower responding. The reversal of the direction of the Mapping main effect between experiments is indicated by the crossover interaction of Mapping and Experiment, $F(1.76) = 100.55$, $p < .001$, $MS_e = 2723.50$. The fact that unlike in the previous experiments, Mapping and Age did not interact, probably has to do with effects of the stimulus material. A complex set of interactions involving Mapping and Experiment indicates that some aspects of the material unintentionally covaried with the Mapping manipulation. Speculations about the explanation of the unexpected effects of Mapping between experiments can be found in the results section of the individual experiments.

[60] In Experiment 5(a), the Stroop effect was only 1.5 ms larger with vocal compared to manual responding.

$F(2,152) = 3.05$, $p = .050$, $MS_e = 182.03$.[61] This interaction is depicted in the upper right panel of Figure 14. The effect of arrow distractors on the Stroop effect was similar in both age groups, whereas old adults were more affected by word distractors than young adults.[62] As can be seen in the Figure, the pattern of *facilitation* effects caused by the different types of Distractor crossed over between Age groups, with word distractors causing larger facilitation for old adults, and arrow distractors causing larger facilitation for young adults. Thus the interaction becomes significant in the facilitation contrast, $F(1,76) = 7.82$, $p = .007$, $MS_e = 374.60$. In contrast, the slight modulation of the age difference in interference by Distractor type fails to reach significance (p=.175).

Because the interaction effect was not further modulated by Mapping or Response modality, it seems safe to conclude that word distractors cause a larger Stroop effect for old adults than for young adults. The reasons for this remain unclear, but one possibility could be that reading might be somewhat more automatic in old adults. Although the power law of learning suggests that increases in speed are only minimal once a skill is sufficiently practiced, the use of extreme groups in a cross-sectional design might have caused an age difference large enough to reveal such minimal effects. An alternative reason could be that many of the old adults came from an academic background, whereas many of the young adults were still high school students. Finally, there might also be cohort differences in the way relative weight assigned to the importance of reading by peers or the society at large.

Having discussed all effects that involved Age, I will now turn to additional consistent or relevant effects that are interesting with regard to the dual-route model.[63] In a later paragraph, the remaining inconsistent effects will be discussed. Vocal responses were faster with arrow than with word distractors, and manual responses were faster with word than with arrow distractors, as indicated by the strong disordinal interaction of Response modality × Distractor type, $F(1,76) = 142.88$, $p < .001$, $MS_e = 1540.81$. However, this is not an effect of stimulus-response compatibility between distractor and response alone, but mainly of stimulus-response compatibility between target and response, as is shown by the consistent interaction of Mapping × Response Modality × Distractor type, $F(1,76) = 104.84$, $p < .001$, $MS_e = 1309.79$. With compatible Mappings, vocal responses were faster with word targets (arrow distractors) than with arrow targets (word distractors), and manual responses were faster with arrow targets (word distractors) than with word targets (arrow distractors). With arbitrary Mappings, both vocal and manual responses were slower with arrow than with word distractors but the effect is somewhat larger for manual responses (17 vs. 8 ms difference between word and arrow distractors, manual vs. vocal). The interpretation of this interaction is made difficult by the fact that the interaction of Response modality and Mapping was highly inconsistent between experiments, see below.

The dual-route model predicts that incongruent high-overlap distractors cause more interference than low-overlap distractors, because they directly prime the incorrect response. Thus large Stroop effects are expected with arrow distractors and manual responses, and with word distractors and vocal responses. The expected crossover interaction of Response modality and Distractor type in the Stroop effect was obtained, $F(2,152) = 65.66$, $p < .001$, $MS_e = 191.91$, and no further modulated by Age (see the lower right panel of Figure 14). The Stroop effect was small (and of similar magnitude) on vocal response trials with arrow distractors and on manual response trials with word distractors (11 and 9 ms, respectively). It was larger on vocal response trials with word distractors and on manual response trials with arrow distractors (32 and 24 ms, respectively). The direction of the crossover interaction was identical between experiments, although the size of the interaction differed between experiments, $F(2,152) = 4.59$, $p = .012$, $MS_e = 191.91$, because the Stroop effect in the manual response condition was smaller in Experiment 5(a). Taken together, the size of the Stroop effect was affected by the degree to which distractor codes had direct access to the response codes. When codes overlapped, the Stroop effect was large, whereas it was small when the distractor had to be translated, e.g. via the semantic route.

---

[61] Disregarding Age, the effect of Distractor type on the Stroop effect was very different between experiments, as indicated by the triple interaction of the two factors with Experiment , $F(2,152) = 9.03$, $p < .001$, $MS_e = 182.03$, whereas globally, Distractor type failed to modulate the Stroop effect, $F(2,152) = 2.62$, $p = .076$, $MS_e = 182.03$, for the interaction of Distractor type and Stroop condition. Word and arrow distractor caused very similar Stroop effects in Experiment 5(b), although the effect was somewhat stronger with arrows. In contrast, the Stroop effect in Experiment 5(a) was larger with word distractors, owing to their larger facilitatory effects.

[62] Additionally, old adults were generally more affected by word than by arrow distractors, whereas young adults tended to be more affected by arrow than by word distractors. However, this latter effect was inconsistent between experiments. Arrow distractors were much weaker in Experiment 5(a), where Stroop effects were generally larger with word than with arrow distractors. In Experiment 5(b), word and arrow distractors were of similar strength globally, but the influence of word distractors on the Stroop effect was very different between age groups.

[63] For the sake of completeness, the two remaining significant, but more or less uninteresting effects are presented here: There was a significant main effect of Stroop condition, $F(2,152) = 161.66$, $p < .001$, $MS_e = 332.55$, which was expected because of the robust nature of the Stroop phenomenon. Mean facilitation was 9 ms, and mean interference was 10 ms. The Stroop effect was somewhat larger in Experiment 5(b), as indicated by the significant interaction with Experiment, $F(2,152) = 3.50$, $p = .033$, $MS_e = 332.55$. The main effect of Distractor type, $F(1,76) =$, $p = .006$, $MS_e = 1911.84$, was significant, but might not be particularly relevant, as its size was only 6 ms overall, and only 1 ms in Experiment 5(b). While it did not change sign between experiments, it was larger at 11 ms in Experiment 5(a), $F(1,76) =$, $p = .023$, $MS_e = 1911.84$.

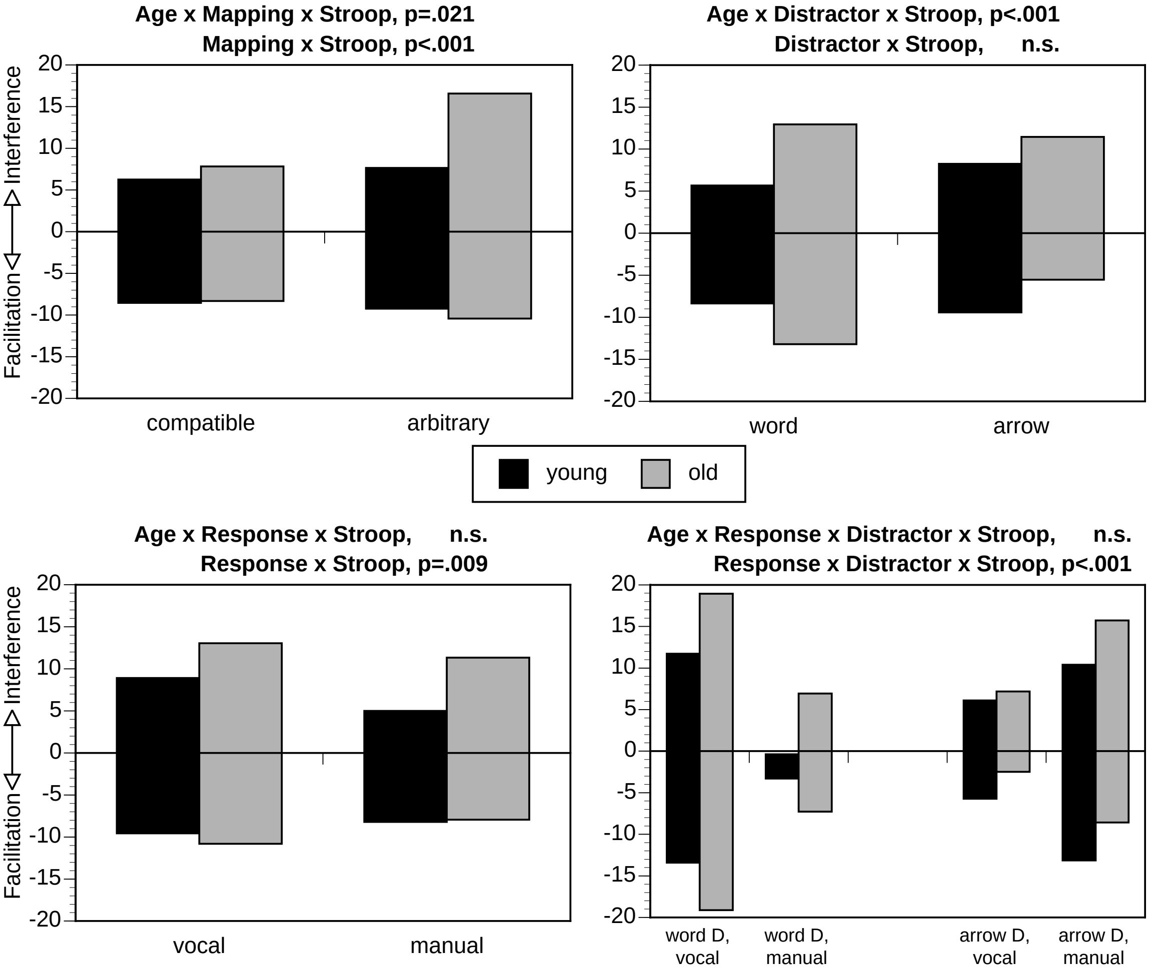


Figure 14. Experiment 5: Interactions of Stroop condition and Age group with other factors in the combined analysis. Bar height corresponds to the Stroop effect. Bars are centered to depict reaction time differences to the neutral condition in incongruent or congruent conditions as deviations from the line at zero ms. Positive values represent interference, and negative values facilitation effects. Panel titles indicate which interaction with Stroop condition is depicted. *Upper left:* An age difference in the Stroop effect is obtained under arbitrary, but not under compatible mapping conditions. This interaction was independent of Response modality (not shown). *Upper right:* Word distractors affect old adults more than young adults, while arrow distractors do not. The size of the facilitation effects caused by different distractors crossed over between age groups. *Lower left:* The age difference in the Stroop effect is equivalent between response modalities. *Lower right:* The very strong interaction of Response modality, Distractor Type, and Stroop condition is independent of Age.

Although the interaction of Age, Mapping, Distractor type, and Stroop condition was far from significant ($F < 1$, $p = .803$) and its direction was the same between experiments, it is nevertheless interesting to compare possible effects of distractor-response compatibility on age differences in Stroop interference within the compatible mappings. Consider what the comparison between high- and low-overlap distractors means: with high-overlap distractors, the distractor directly activates a response code, while the target code needs to be translated. With low-overlap distractors, the target directly activates the response, and the distractor needs to be translated. However, in any case, with the compatible mappings the translation of distractor or target into a response is automatic because of highly overlearned associations between arrows or words and directional concepts on the one hand, and between directional concepts and directional responses on the other hand.[64] Thus although a translation is required, in contrast to the arbitrary mappings this translation does not require controlled processing along the episodic route. Hence, a comparison of the effects of high- and low-overlap distractors in the compatible mapping conditions allows to evaluate the extent to which a translational process is effected by aging, when this process does not take the episodic route. If translation per se is affected by aging, then larger age

---

[64] The strength of association of a verbal direction word and a spatial concept, and between a spatial concept and a verbal response code might be learned relatively late in life (Clark, 1973; Levinson, 2003), however, it is at least relatively much more established than the arbitrary associations that were built up during the experiment.

differences in the Stroop effect are expected with high-overlap distractors. Possible age differences in code translation were evaluated with an analysis of variance limited to the compatible conditions. The factors included Response modality, Distractor-response compatibility, and Stroop condition. To obtain the DRC factor, levels of Distractor type were rearranged depending on response modality. In this analysis, the interaction of Age, DRC and Stroop condition was far from significant, $F(2,152) < 1, p = .799$, and consistent between experiments. From this result one can indirectly infer that the critical Age×Mapping×Stroop condition interaction is due to working memory demands, not translation per se.

Finally, let us turn towards additional results that were highly inconsistent between experiments. These include the interaction of Mapping and Response modality, $F(1,76) = 55.14$, $p < .001$, $MS_e = 1175.61$, and the interaction of the two factors with Stroop condition, which failed to reach significance on an overall level, $F < 1$, but interacted with Experiment, $F(2,152) = 5.62$, $p = .004$, $MS_e = 159.34$. Overall, the Mapping effect was somewhat larger with vocal than with manual responses. However, there was a disordinal Mapping × Response modality × Experiment interaction, $F(1,76) = 16.34$, $p < .001$, $MS_e = 1175.61$. In Experiment 5(a), the Mapping effect was in the standard direction for both response types (38 and 27 ms, vocal and manual). In Experiment 5(b), a very small Mapping effect in the standard direction was observed with vocal responses (2 ms), but a reversed Mapping effect with manual responses (-34 ms). As has been discussed in the results section of Experiment 5(b), this pattern is probably an artifact of the choice of stimuli. In both experiments, Stroop interference was larger with arbitrary than with compatible mappings in both response modalities. However, the direction of the mediating influence of Response modality on the Mapping effect in Stroop interference differed between Experiments. In Experiment 5(a), the Mapping effect (in Stroop interference) was larger for manual responses (and nearly absent for vocal responses). In Experiment 5(b), the Mapping effect (in Stroop interference) was larger for vocal responses (and nearly absent for manual responses). These inconsistent effects are due to problems with the stimulus material. It should be clear that no firm conclusions can be drawn about higher-order effects that include combinations of Mapping and Response modality.

Proportional measures. To not unnecessarily lengthen the results section, only age-related effects of the analyses using proportional measures are reported. In the analysis of log reaction times following the same factorial pattern as the analysis of untransformed reaction times, three effects involving Age were significant, namely the main effect of Age, $F(1,76) = 61.12$, $p < .001$, $MS_e = 6.41\text{e-}2$ and the interactions of Age and Response modality, $F(1,76) = 24.39$, $p < .001$, $MS_e = 3.38\text{e-}2$, and of Age, Distractor type, and Stroop condition, $F(2,152) = 7.83$, $p = .001$, $MS_e = 6.95\text{e-}4$. The direction and interpretation of these effects is the same as in the untransformed RT analysis.

The critical interaction of Age, Mapping, and Stroop condition was only marginally significant, $F(2,152) = 2.47$, $p = .088$, $MS_e = 7.44\text{e-}4$, which compared with the untransformed analyses might indicate that part of the effect can be explained by general slowing. However, a lack of significance for the interaction might also be caused by the fact that the Stroop effect is numerically larger for old adults with both Mappings, (although the age difference is not large enough to cause an interaction of Age and Stroop condition), albeit larger with the arbitrary mapping (although the modulating influence of Mapping is not large enough to cause the triple interaction to become significant). In separate analyses performed for each level of Mapping, Age and Stroop condition interacted significantly for the arbitrary Mappings, $F(2,152) = 3.98$, $p = .021$, $MS_e = 1.11\text{e-}3$, while the interaction was far from significant for the compatible Mappings, $F(2,152) < 1, p = .913$. Thus at least for arbitrary mappings, the age difference in the Stroop effect appears to be over-proportionally large.

In the analysis using the proportional Stroop effect measure $(incongruent\ RT - congruent\ RT)/neutral\ RT$, the interaction of Age and Mapping was marginal, $F(1,76) = 3.23$, $p = .076$, $MS_e = 1.88\text{e-}3$. However, if a different measure was used, namely the proportional *interference* measure $(incongruent\ RT - neutral\ RT)/neutral\ RT$, the interaction of Age and Mapping was significant, $F(1,76) = 4.02$, $p = .048$, $MS_e = 2.09\text{e-}3$. Again, if the analysis was limited to the arbitrary mapping conditions, the Age main effect was significant, $F(1,76) = 6.79$, $p = .011$, $MS_e = 3.03\text{e-}3$, while it was far from significant when only compatible mapping conditions were analyzed, $F(1,76) < 1, p = .722$. Because a relative interference measure was analyzed, which already accounts for age differences in baseline responding, this means that age difference in Stroop interference were over-proportional in the arbitrary mapping conditions, while there were no age differences in (proportional) Stroop interference in the compatible mapping conditions.

The interaction of Age and Distractor type was only significant in the proportional Stroop effect measure, $F(1,76) = 17.93$, $p < .001$, $MS_e = 1.61\text{e-}3$, but not in the proportional interference measure, $F(1,76) = 1.95$, $p = .166$, $MS_e = 1.74\text{e-}3$. The crossover pattern thus appears to be caused by age differences in facilitation: congruent word distractors caused relatively large facilitation effects for old adults, and congruent arrow distractors caused relatively large facilitation effects for young adults.[65]

---

[65] Facilitation has sometimes been interpreted to reflect an automatic component of the Stroop effect (Posner & Snyder, 1975). According to this interpretation, one cause for the pattern could be cohort differences in the learning history with the two stimulus classes. Old adults have a longer reading history, while for young adults reacting towards spatial stimuli might feel more natural, because they have had more exposure to

None of the effects reported in this section was further modulated by an interaction with the Experiment factor, thus the effects can be considered consistent between experiments.

Errors. Analogous to the procedure in the reaction time analysis, data from both experiments were pooled and Experiment was included as an additional between subjects factor in the combined analysis of variance.

In the error analysis, four main effects were significant, namely Age, $F(1,76) = 17.24$, $p < .001$, $MS_e =$ 1.64e-3, Mapping, $F(1,76) = 9.80$, $p = .002$, $MS_e =$ 1.93e-3, Stroop condition, $F(2,152) = 95.78$, $p <$ .001, $MS_e =$ 6.0e-4, and Response modality, $F(1,76) =$ 5.54, $p = .021$, $MS_e =$ 4.20e-3. Young adults made more errors than old adults (*Ms* =3.79 vs. 2.44 %), there were more errors with the arbitrary than with the compatible Mapping (*Ms* =2.80 vs. 3.44 %), and errors due to Stroop condition were in the expected order, with least errors with congruent, intermediate with neutral, and most with incongruent distractors (*Ms* =2.31, 2.85, and 4.19 %). Interference was larger than facilitation. More errors were committed with manual than with vocal responses (*Ms* =3.47 vs. 2.76 %), but the Response modality main effect was qualified by the Age × Response modality interaction, $F(1,76) = 11.37$, $p =$ .001, $MS_e =$ 4.20e-3. Only young adults made more errors with manual than with vocal responding, while there was no effect of response modality for old adults—numerically, they even made fewer errors in the manual modality. Like in most other experiments, with manual responses there appear to be age-differential speed-accuracy criterion settings, with young adults accepting a higher error rate to achieve fast responding.

Even though the critical interaction of Age, Mapping, and Stroop condition failed to reach significance, $F(2,152) = 2.52$, $p = .084$, $MS_e =$ 3.69e-4, the direction of the effect was the same as in reaction times. Young adults' Stroop effects in errors tended to be smaller, while old adults' Stroop effects were larger with arbitrary than with compatible Mappings. Thus the pattern in errors, although not significant, supports the conclusion drawn from the reaction time results, that working memory requirements induced by arbitrary mapping rules amplify the Stroop effect in the group of old adults.

The critical interaction of Age, Response modality, and Stroop condition that evaluates whether old adults might be particularly affected by distractors when responding manually also failed to reach significance, $F(2,152) = 1.90$, $p = .153$, $MS_e =$ 5.09e-4. Furthermore, there was a marginal interaction with Experiment, $F(2,152) = 2.65$, $p = .074$, $MS_e =$ 5.09e-4. Overall, if anything, there was a tendency for a *larger* Stroop effect for young than for old adults in the manual response condition, while there was absolutely no Age difference in the Stroop effect for vocal responses. However, the Age difference in the manual modality was only obtained in Experiment 5(b) and is thus inconsistent between experiments. With regard to the research question, both the reaction time and the error results indicate that there is no particular interference-proneness in the manual modality for old adults.

A pattern of interactions that was consistent between experiments involved Response modality and Distractor type. The two factors interacted, $F(1,76) = 13.79$, $p <$ .001, $MS_e =$ 2.31e-3, and were involved in a triple interaction with Stroop condition, $F(2,152) = 34.79$, $p <$ .001, $MS_e =$ 4.92e-4, and in the four-way interaction of Age, Response modality, Distractor type, and Stroop condition, $F(2,152) = 3.13$, $p = .047$, $MS_e =$ 4.92e-4. There were no differences between response modalities with word distractors, while there were fewer errors with vocal responding and more errors with manual responding when the distractor was an arrow. As expected, a crossover interaction between Response modality and Distractor type was obtained with regard to the Stroop effect. Word distractors caused more errors than arrow distractors with vocal responses, and the reverse was true for manual responses. The triple interaction reduces to a double interaction if Distractor type is recoded into overlap of distractor and response codes (or distractor-response compatibility), contingent on Response modality. While there was a larger Stroop effect with high-overlap distractors than with low-overlap distractors, the error Stroop effect did not differ between response modalities for either high-overlap distractors (3.0 vs. 2.8 percent points difference, vocal vs. manual), or low-overlap distractors (0.6 vs. 1.0 percent points difference, vocal vs. manual). These results fully parallel the reaction time results and indicate a contribution of the direct route to the Stroop effect. The result is compatible with a dual-route model, because in addition to distractor-response compatibility, the manipulation of which is similar at both levels of Mapping, the manipulation of Distractor type at the compatible Mapping also meant a change in target-response compatibility. Because response-compatible targets lead to a fast activation of the response, and distractors in these conditions have to be processed via the indirect route, the chance for a distractor to influence the response is relatively low in conditions with compatible mappings and low-overlap distractors. Thus the incongruent-congruent difference in error rate was close to zero in the low-overlap conditions. The Stroop effect was higher whenever translation of the target code into the response code was required. This was the case with arbitrary mappings as

---

computers, video games and the like. Furthermore, translation of the words left and right into spatial codes is acquired rather late in life, which might make them relatively weak distractors for high school students. For example, there is evidence from developmental psychology that the concepts left and right are acquired later than under and top (Clark, 1973), possibly due to a lack of supportive body asymmetries for left and right. Furthermore, even 4-year-old children have trouble performing left-right discriminations: "Western children learn topological spatial terms first, starting at about age 2, proceed to intrinsic uses, and about the age of 4 have relative usages of front and back; but left and right terms lag far behind, with relative left and right often not being fully mastered before 11."(Levinson, 2003).

well as with compatible mappings when there was high overlap between distractor and response codes.

The four-way interaction was only obtained in errors, not in reaction times. In the vocal response modality, the difference in the Stroop effect when comparing word and arrow distractors was identical between age groups. In the manual modality, the difference between the relatively large Stroop effects caused by arrows and the relatively small Stroop effects caused by words was larger for young than for old adults. Like in reaction time, arrow distractors appear to have a stronger influence on young adults' Stroop effects. Unlike in reaction times, the strength of the interaction was modulated by Response modality, with larger age differences occurring with manual responses. However, the direction of the effect indicates that young adults, not old adults might have a problem ignoring irrelevant arrows.

The last effect involving age group was the interaction of Age, Distractor type, and Stroop condition, $F(2,152) = 3.94$, $p = .021$, $MS_e = 4.48\text{e-}4$. Although the interaction was further modulated by Experiment, $F(2,152) = 4.39$, $p = .014$, $MS_e = 4.48\text{e-}4$, there were consistently larger Stroop effects with arrow distractors for young than for old adults. With word distractors, either no Age difference in the Stroop effect (Experiment 5(a)), or larger Stroop effects for old than for young adults (Experiment 5(b)) were obtained. Results thus parallel the reaction time pattern indicating that either arrow distractors are particular disturbing for young adults, or word distractors are particular disturbing for old adults, or both. Distractor type and Stroop condition also interacted on an overall level, $F(2,152) = 3.86$, $p = .023$, $MS_e = 4.48\text{e-}4$, with word distractors causing somewhat larger Stroop effects. The interpretation of the present interaction is complicated by a strong and inconsistent interaction of Distractor type, Stroop condition, and Experiment, $F(2,152) = 7.03$, $p = .001$, $MS_e = 4.48\text{e-}4$, which indicates that the influence of Distractor type on the Stroop effect was inconsistent between experiments. While in Experiment 5(a), the Stroop effect was larger with word than with arrow distractors, the pattern was reversed in Experiment 5(b).

Finally, one consistent effect that was equivalent between age groups was the interaction of Mapping, Response modality, Distractor type, and Stroop condition, $F(2,152) = 8.41, p < .001, MS_e = 3.88\text{e-}4$. A crossover interaction between Response modality and Distractor type in the Stroop effect was observed with both Mappings, but the size of the interaction was larger with the compatible than with the arbitrary mapping. With compatible mappings, the incongruent-congruent difference in error rate was .0028 with vocal responses and word/arrow ensembles, and .0032 with manual responses and arrow/word ensembles, and it was .032 with both ensembles featuring high overlap between distractor and response codes. With arbitrary mappings, the difference in the effect of low- and high-overlap distractors was less pronounced. This latter fact had not been observed in the reaction time analysis, where the interaction of Response modality, Distractor type, and Stroop condition was not modulated by Mapping. The fact that Stroop effects due to high-overlap distractors were even larger in the compatible than in the arbitrary mapping might indicate that response priming decays relatively rapidly, so that the activation of the response code following the slow table-lookup process needed in the arbitrary mappings arrives at a point in time where little distractor-based priming is left. Alternatively, inhibition of the wrong response at the level of episodic accumulators could have spread to the response module. Part of the four-way interaction is also caused by the fact that there was no difference in the size of the Distractor type × Stroop condition interaction with the compatible mappings, but with the arbitrary mappings, the Stroop effect was somewhat more affected by distractor type with vocal than with manual responding. The reason for this is unclear.

Lastly, let me mention one effect that was fairly inconsistent between experiments, namely the interaction of Mapping and Response modality, $F(1,76) = 18.41$, $p < .001$, $MS_e = 2.17\text{e-}3$, which was further modulated by Experiment, $F(1,76) = 8.47$, $p = .005$, $MS_e = 2.17\text{e-}3$. The Mapping effect was generally stronger for vocal than for manual responding. A standard Mapping effect, i.e., less errors with a compatible than with an arbitrary Mapping, was only obtained in the vocal modality. However, the fact that no Mapping effect at all was obtained in the manual modality was due to inconsistencies between experiments that are likely caused by problems with the choice of stimuli. In Experiment 5(a), the tendency was in the standard direction, while in Experiment 5(b), a reverse mapping effect was observed for errors like for reaction times in the manual response condition.

In summary, error and reaction time results were fairly consistent, the primary exception being the age-differential speed-accuracy tradeoff with manual responding.

Discussion

There were problems with the stimulus material in both individual experiments. In Experiment 5(a), where spatially separated stimulus ensembles were used, the target could be detected by using information about differences in spatial frequency between target and distractor in all but one stimulus condition, which used shape-arrow ensembles. This caused reaction times to be particularly slow with that ensemble. In Experiment 5(b), where target and distractor were displayed at the same spatial location, a different problem was encountered, which might be related to figure-ground perception. Responding to a word that was printed in background color inside of an arrow printed in foreground color was relatively slow. In contrast, responding to the stimulus quality color in the arbitrary condition was particularly fast, possibly because the color was part of the foreground, and additionally, the arrow was rather large. Taken together, this led to a reverse Mapping effect for stimuli featuring arrow distractors.

Despite the fact that unintentional side effects of the material led to a more complex pattern of results than intended, results of the combined analysis show that whether or not old adults are affected by larger Stroop interference than young adults depends on the type of relation between stimulus set and response set. When the target dimension can activate the response concept without mediation of memory processes, as in the compatible mapping conditions, there is no age difference in interference. This lack of an age difference is observed independently of whether the target code requires to be translated into a response code, or whether it can directly activate a response. In contrast, with arbitrary mapping conditions, where translation of a target into a response code needs controlled processing along the episodic route, because it relies on relatively recently acquired associations, an age difference appears.

The fact that the critical interaction effect only reached significance in the combined analysis, while there was only a tendency in both single Experiments 5(a) and 5(b) does not invalidate the argument, because the direction of the effect was the same in both experiments. In a way, the current experiments provided an extreme test of the reliability of mental sets hypothesis, because (a) the arbitrary mapping only required to hold on-line two arbitrary rules concerning stimulus-response pairings[66], (b) within the experiment, twice as many trials were available for a given target stimulus with the arbitrary than with the compatible mapping, and (c) target and distractor switched their roles in the same stimulus configuration (arrows/words) only for the compatible mappings, which might have introduced higher executive demands in the compatible conditions. Hence the memory demands in the arbitrary conditions were close to the smallest size imaginable. Thus results provide converging evidence that the age effect in interference is larger when demands on the episodic route (or working memory, or the active part of long-term memory) are increased, even if the increase is conceivably small.

This effect is relatively independent of response modality. Although like in the previous experiments, the age difference in response times was larger with manual than with vocal responses, there was no sign of an age-differential modulation by response modality of early difficulty effects, as measured by the Stroop effect. With the appropriate caution regarding the interpretation of null effects, it can be concluded that response modality per se does not modulate the age difference in the Stroop effect. If anything, young adults tended to produce larger Stroop effects than old adults when responding manually. This stands in stark contrast to results from experiments that confounded Response modality and Mapping, where old adults produced larger Stroop effects, and the age difference in the Stroop effect was particularly large with manual responses. Therefore, the interpretation of the age-related manual-response results from the earlier experiments as Mapping effects receive support from the present results. The conjecture that the previous results obtained were merely due to the fact that the arbitrary mapping was implemented by using manual responses can be rejected.

The logic of the current experiment assumed that manual responses rely on spatial codes and vocal responses on phonological codes, and current models of word production assume that phonological codes have privileged access to lexical representations. Although the consistently observed interaction of Response modality and Age (independent of Stroop condition), with larger age effects in the manual than in the vocal modality, might partly be explained by a different criterion setting of the speed-accuracy tradeoff, it could also indicate that the type of response code is a mediator of age effects. It could therefore be related to the lexical/nonlexical distinction that has been brought forward by Myerson et al. (1994) and is supported by previously published meta-analytic results. The latter view is also supported by another result from the current experiments: age differences in the Stroop effect were larger with word distractors than with arrow distractors. Although this effect was somewhat unexpected, it happens to be compatible with Myerson et al.'s distinction, if it is assumed that an arrow is primarily processed along a nonlexical route, while a word automatically activates the corresponding lexical entry. In this case, stimulus codes, not response codes are relevant. In cases where the direct route is available, i.e., in high-DRC conditions, distractors might directly activate the corresponding response code. In low-DRC conditions, processing of the distractor might still be automatic (because the distractors' meaning was always overlearned), but a lexical or conceptual code is activated instead of a response code. Apparently, the degree to which arrows and words caused automatic activation was different between age groups, as the interaction of Age, Distractors type, and Stroop condition was partly due to age group differences in the facilitatory influence of the distractors. Regardless of response modality (and hence of DRC), old adults' responses were less facilitated by congruent arrow distractors than young adults'. Also independent of response modality, old adults experienced more facilitation from word distractors than young adults. Additionally, while the size of the Stroop effect caused by arrow distractors did not differ between age groups, old adults were always more disturbed by word distractors than young adults, which might indicate a stronger reliance on lexical codes in the old group.

One interpretation of the distractor type influence is that old adults are particularly susceptible to interference from phonological codes. An alternative explanation is that young adults are relatively more susceptible to interference from spatial codes. Speculations about a possible age-differential learning history were already brought forward in the discussion of Experiment 5(b). Because the representation of the mapping rules must also be coded in some way, whether lex-

---

[66] In fact, researchers investigating the effects of arbitrary mappings often recommend to use at least three-element sets, because with two-element sets the response on a response change trial can be found by simple alternation.

ically or nonlexically, a good way to test this would be to examine whether there are interactions involving Distractor type, Mapping, and Age. If old adults tend to rely on phonological coding to represent arbitrary mappings, while young adults choose a more flexible strategy—e.g. choose a phonological code for vocal responding, and a visuo-spatial code for manual responding, then the influence of Distractor type on the Mapping effect should differ between Age groups (in the example, the interaction should also be modulated by response modality). Unfortunately, the interaction of Mapping and Distractor type was highly inconsistent between experiments, which circumvents a reasonable interpretation of the relevant higher-order interactions.

This inconsistent interaction also prevents the conclusion that the distinction between the lexical and the nonlexical domain cannot explain the observed age-related modulation of Stroop interference by Mapping. Although that interaction was neither modulated by Response modality nor by Distractor type, more research is needed to reveal whether phonologically coded distractors affect old adults' maintenance of arbitrary rules more than other, e.g., spatial distractors.

A null results that was consistent between Experiments 5(a) and (b) is also relevant for the discussion of age differences in the Stroop effect. While there was a strong and consistent interaction of Response modality, Distractor type, and Stroop condition, the size of this interaction was (consistently) similar in the two age groups. The former interaction supports a dual-route model of the Stroop effect, and is easier to describe if distractor type is recoded into distractor-response compatibility: with high overlap of distractor and response codes, distractors can prime the associated response, which leads to large Stroop effects. In parallel, the associated distractor semantics are activated along the indirect route. With low distractor-response overlap, the direct route is unavailable, so that interference can exclusively arise along the indirect route, which leads to much reduced, albeit still significant Stroop effects. The lack of an interaction with Age indicates that the strength of the automatic activation of either a response code or a conceptual code by the distractor does not differ between age groups.

In summary, age differences in the Stroop effect are more pronounced (a) when the mappings of target concept to response is arbitrary, and possibly (b) when verbal as opposed to spatial distractors are used. In contrast, they are not affected by response modality or by the degree of automatic activation of a response set member (i.e., direct stimulus-response compatibility). This strongly suggests a specific age deficit in working memory. Even reaction time tasks that—if compared to working memory tasks proper—appear to be relatively undemanding often rely on working memory processes to some degree. These rather specific demands of reaction time tasks selectively amplify difficulty effects resulting from earlier processing. Because these demands on controlled processing often scale with task complexity, they might be a specific factor responsible for the Brinley plot pattern.

Experiment 6

Stroop task with word-list versus color-matching instructions

Considering the research program conducted by Hale, Myerson, and collaborators, one alternative explanation might account for at least part of the pattern of results observed so far. This group has consistently found that old adults are much more impaired in visuospatial than in verbal tasks, with the spatial-verbal partition roughly corresponding to the similar partition implemented in the Baddeley/Hitch model of working memory. For example, in two studies comparing visuospatial/nonlexical and verbal/lexical tasks (Jenkins et al., 2000; Hale & Myerson, 1996), Brinley slopes for the former were rather large (2.56 and 3.11, respectively), while Brinley slopes for the latter were close to one (1.22 and 1.35).

If old adults introspectively know about their difficulties with spatial tasks, they might prefer to use a verbal strategy whenever there is an option. For example, if they are told to respond to a blue color patch with the left key, they might use verbal working memory to code the rule BLUE→LEFT. Young adults, on the other hand, might prefer to rely mainly on spatial working memory, using a representation resembling a spatial layout of colors. A manual-response in a Stroop-like task with an arbitrary concept-to-response mapping could pose particular problems for old adults, because (a) a spatial code is used for responding, and (b) a word distractor interferes with the phonological code that is used for storage and maintenance processes in verbal working memory.

The result from the last experiment that old adults are more affected by word distractors, while there is not much of an age difference in interference due to arrow distractors could also be related to the spatial/verbal distinction. For the arbitrary mapping conditions, it could indicate that young adults tend to rely on spatial working memory, whereas old adults due to their documented difficulties with spatial material might prefer to use verbal mechanisms. A second process, age-differential strength of automatic translation processes along a spatial and a verbal route, might also have contributed to the distractor type effect with the arbitrary mappings, and might have been the sole reason for the effect with the compatible mappings. The relative speed of activation of spatial response codes by an arrow and by a word might differ between age groups: for young adults, arrows are translated into spatial response codes very fast, which protects arrow targets from word interference, but causes arrow distractors to interfere with the translation of a word into a spatial response code. For old adults, the activation of a spatial response code might generally take more time, thereby making arrows weaker distractors. Old adults, however, may be particularly efficient readers, i.e., the translation of a written word into a phonological response code is as fast as for young adults. This makes words relatively strong distractors, especially if the activation of spatial response codes takes relatively long (an indicator of which the consistently observed pattern of larger age effects with manual responses might be).

Whereas there might be differences in the speed of activation of automatic translation processes, the present experiment focusses on possible age differences in verbal and visuo-spatial working memory. The experimental strategy was to use (a) tasks that differ in the degree to which they are amenable to a spatial or to a verbal representation, namely a Stroop task and a reverse Stroop task, and (b) to instruct different groups of participants to use either a verbally or a spatially coded task set. Because controlled memory processes were at the focus of investigation, the tasks always involved an arbitrary mapping of target concepts to manual responses.

By definition, stimuli in a manual-response task with an arbitrary mapping never have a pre-established spatial association. Because there is no natural or overlearnt relationship between colors and locations, the task has a strong episodic component. If a manual response is to be given to a stimulus with color semantics, e.g. a color patch or a color word, then the perceptual code has to be translated into a spatial response code. Not only do the codes have to be translated, but translation is a rather controlled process, i.e. the matching response set element has to be found by "table lookup". How does this table lookup work? Consider this (a) for a rule set maintained in a spatial medium (using an analogue, visuo-spatial code), and (b) for a rule set maintained in a verbal medium (using a phonological code).

In the former case, the representation of the color-to-key rules in the visuo-spatial sketchpad could be akin to a spatially arranged layout of colors. Thus spatial working memory would connect perceptual color codes with spatial response codes. On the other hand, maintenance in the verbal subsystem will rely exclusively on phonological codes. The representation will therefore resemble a word list, where the phonological representation of the color concept is associated with a phonological representation of a spatial response concept.

On the input side, different types of stimuli differ in the degree to which they afford access to verbal or spatial working memory. For example, a color patch might be well-suited as input into *spatial* working memory, and similarly, a color word might be better-suited as input into *verbal* working memory. In both cases, one part of the association is directly activated by the stimulus, without a detour via the conceptual level—the stimulus code is compatible (overlaps) with the working memory code. On the other hand, a word must be translated into a color via the conceptual route to be used as input into spatial working memory. Likewise, a color percept can only reach verbal working memory by activation of the corresponding phonological code via the conceptual route.

On the output side, manual responses seem to be generally better suited to mappings stored in *spatial* working memory. In this case, the spatial code of the entry that is retrieved by episodic "table lookup" can be directly read out by the response system. With reliance on verbal working memory, the result of the controlled lookup process is a phonological code (e.g. `/top/`), which needs to be translated into the spatial code used by the response system. Although this latter translation process is overlearnt, it will still take some time. Hence, manual responding should generally profit from a spatial coding of arbitrary S-R rules. The layout of the response keys is spatial, and thus the output of spatial working memory can directly activate a topologically organized response code, while verbal working memory output has to be translated.

Thus if manual responding to the meaning of a *color word* is required, some higher-level translation needs to be performed, regardless of the working memory compartment used. If the task set is mainly stored in verbal working memory, then the phonological rule output must be translated into a spatial response code. Conversely, if the task set is stored in spatial working memory, then at the input side the graphemical code must be translated into a perceptual color code amenable to an analogue, spatial representation. In contrast, with manual responding to a *color patch*, translation with respect to working memory is not required if the spatial store is used: responding should be relatively fast, because the perceptual code directly matches the stored representation, and the output code directly matches a response code. On the other hand, double translation is required if the verbal store is used, in which case the code needs to be translated at input and output.

While the involved translation processes are rather automatic, they nevertheless take time, presumably due to spreading of activation between relatively distributed systems. This is the reason why spatial coding of arbitrary color-to-key rules should be the preferential strategy in a task that requires manual responding to colors.[67] In comparison to individuals who prefer to use a verbal code, spatial 'coders' should show a response time benefit in a task that uses color stimuli (e.g., the neutral condition of a Stroop task). They should also be faster in that task than in a task that uses verbal stimuli (e.g., the neutral condition of a reverse Stroop task), for which the differences between spatial and verbal coding strategies should be less pronounced, because recoding is required in either case.

Let us now turn towards multivalent color-word stimuli. Consider the influence of distractors on working memory: Word distractors (in the standard Stroop task) will have premium access to the verbal compartment of working memory, and will interfere with a mapping rule maintenance strategy that uses a phonological code. Color distractors (in a reverse Stroop task) will have easier access to spatial working memory, and will hence

---

[67] However, note that the benefit of spatial coding might not be very large: the main difference is that the additional, but automatic translation process that is inserted with verbal coding is not required.

cause stronger interference for individuals who rely on a spatial coding strategy for the arbitrary mapping rules.

Thus in the manual-response *Stroop task*, a spatial coding strategy would pay double, (a) because it is most appropriate for the association of stimulus and response code, and because interference by phonologically coded distractors occurs only indirectly, via the semantic route. In the manual-response *reverse Stroop task*, a phonological coding strategy appears to be more appropriate, despite the translation required at the output end between the phonological code of the spatial concept and the actual spatial response code. Here, the target can activate verbal working memory relatively directly, and colors have less chance to interfere with verbal than with spatial working memory.

It could be that old adults, who are 'spatially impaired', do not rely on spatial coding to the same extent as young adults, even when a spatial coding strategy would be most appropriate, e.g., in a color-word Stroop task with manual responding. Old adults might prefer to use an alternative code, possibly mediated by verbal labels. The fact that the age difference in Stroop interference is larger with manual than with verbal responding would then be a consequence of the differential coding strategies, since 'irrelevant' words have a higher chance to interfere with a verbally maintained map than with a spatially maintained one. Results from the previous experiment, indicating that word distractors lead to a larger age difference in the Stroop effects than arrow distractors are consistent with this conjecture.

In the present experiment, it was tried to induce either spatial or verbal coding strategies for the color-to-key mappings by varying the instruction. In the instruction and the mapping acquisition phases, mapping rules were presented in either spatial or verbal layout. The 'spatial' group saw a spatial layout of color patches corresponding to the layout of response keys, while the 'verbal' group saw a list of color-name→direction-word pairs (see Figure 15). Subjects were told that "imagining the spatial layout of colors" or "silently repeating the list of associations" before each trial would aid task performance. It was hoped that the instruction manipulation made participants prefer one representation over the other, thereby inducing either an analogue, spatial representation or a phonological, verbal representation of the rules. The external representation of the mapping rules was visible during all of the mapping acquisition trials, and was also presented during the experiment whenever an error had been made.

A reverse Stroop condition was added to the experiment, using a color word target which was printed in a to-be-ignored color. In addition to the evaluation of the instruction manipulation, the reverse Stroop condition is interesting on its own, because little is known about whether the reverse Stroop effect is affected by aging when arbitrary task rules have to be held on-line. Furthermore, the comparison of Stroop and reverse Stroop effects in a situation in which the response system always uses a spatial code, and hence there is no pre-established association between either words or colors and responses, can provide additional insight in the system used to store the mapping rules.

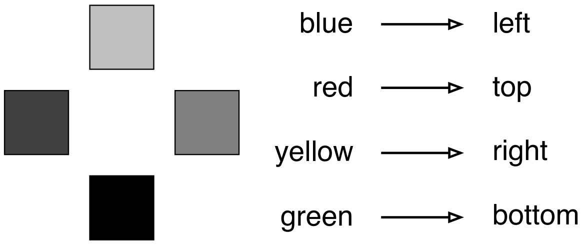


Figure 15. Stimuli used by the instruction manipulation to induce 'spatial' (*left*) and 'verbal' (*right*) representations of arbitrary color-concept to key mapping rules.

Reverse Stroop effects are typically small to nonexistent when the standard, vocal response format is used. Reading is well-practiced, and there is large overlap between the graphemical representation of the word on the input side and the phonological code used for responding, and only little overlap between the color distractor and the verbal response. This leads to a fast activation of the appropriate reading response, and the slower spreading of activation from the color concept to its verbal/lexical representation caused by the ink color has little chance to interfere.

However, if manual responding is used, the reverse Stroop effect should become larger, because the degree of overlap between graphemical target representation and manual response is much smaller. The graphemical input code has to be translated into a spatial response code along the conceptual route, thus giving the color more time to interfere. If this translation makes use of spatial working memory, then interference effects could actually become rather large, because of the degree of match between distractor and memory code. In fact, recent results show that a slight modification of the task leads to a large reverse Stroop effect. This is the case when manual pointing responses (with a computer mouse) towards spatially arranged patches of color are. Using this 'color-matching' Stroop task, Durgin (2000) was able to completely reverse the typical pattern—he observed a very small Stroop effect, but a large reverse Stroop effect. Incongruent word distractors interfered little with reactions towards ink color, whereas incongruent color patches strongly interfered with reactions towards word meaning. This constitutes strong evidence for a contribution of response compatibility to the Stroop effect.

Predictions regarding the instruction manipulation in the current experiment are rather straightforward for the color (Stroop) task. First, interference effects should be larger with the 'verbal' than with the 'spatial' instruction, since the former matches the distractors, while the latter matches the target, which ideally should lead to results similar to the ones obtained by Durgin (2000). I also expected an interaction of Instruction, Age, and Stroop interference. If old adults tend to rely on a verbal coding of mapping rules, then explicitly instructing them to use a more appropriate spatial coding should aid performance. It should aid old adults more than young adults because the latter presumably choose a spatial representation of arbitrary rules whenever the task context affords it. Because a spatial representation is less susceptible to interference from word distractors, the age difference in interference should be smaller with the 'spatial' than with the 'verbal' instruction. Predictions for the word (reverse Stroop) task are less clear because it is not entirely clear how strong the 'lexical' component of internal color representations is. However, in the word task the input side of the mapping rule matches the rules given by the 'verbal' instruction, while the 'spatial' instruction which gives a layout of colors does not match the target dimension at all, but only the distractor dimension. Thus in the word task, a 'spatial' instruction should lead to larger interference effects than a verbal instruction. There are no specific predictions whether this effect should be modulated by age.

Method

Participants. Participants were 40 younger (age $M =$ 20.7 years, $range = 18 - 31$) and 40 older adults ($M =$ 71.6, $range = 66 - 79$), drawn from the University of Potsdam participant pool. The current experiment was performed during a control session of a larger training study (Junker & Kliegl, unpublished), which required participants to come to the lab for four sessions lasting about one hour each. Participants were paid 15 DM (about $ 7) per session for their participation in the study. Age groups did not differ with respect to years of formal education (young, $M = 12.3$, $SD = 2.0$; old, $M = 12.6$, $SD = 3.5$), $t(78) < 1$. With regard to the measures of fluid and crystallized intelligence, the sample was fairly standard: On the Digit Symbol Substitution Test, younger adults outperformed older adults (young, $M = 61.0$, $SD = 8.8$; old, $M = 47.5$, $SD = 7.8$ points), $t(78) = 7.30$, $p < .001$, whereas older adults' performance in the MWT-A vocabulary test was better than younger adults' (young, $M = 31.1$, $SD = 2.4$; old, $M = 32.6$, $SD = 1.9$ points), $t(78) = -3.02$, $p = .003$. All participants were healthy according to self-rating and had normal or corrected-to-normal vision.

Design, Stimuli and Procedure. The design was a 2 × 2 × 2 × 3 mixed factorial, with Age group (young vs. old) and Instruction (spatial vs. verbal) as between subjects variables, and Task (color vs. word) and Stroop condition (congruent, neutral, incongruent) as within subjects variables. The tasks used were Stroop-like tasks with manual responding to color-word stimuli. The within-subjects factor Task required subjects to respond either to word color (Stroop task) or to word meaning (reverse Stroop task), subsequently referred to as Color and Word, respectively. The order of the two tasks was counterbalanced across participants, with matched orders used between instruction conditions and age groups. Stroop condition was randomized within task blocks. The proportion of congruent, neutral, and incongruent trials was .40, .40, and .20 respectively. [68]

Because the instruction manipulation was crucial for the establishment of either a 'spatial' or a 'verbal' set, care was taken to create a vivid representation of the mapping rules. The instruction manipulation involved two steps. First, subjects were told that their performance would benefit if they used a certain strategy of memorizing the color-concept-to-key mapping rules. In the spatial 'color-matching' condition, participants were shown a spatial arrangement of colored patches corresponding to the locations of the associated response keys (see Figure 15, left). They were told that trying to memorize a visual representation of that pattern would be the best way to solve the task. In the 'verbal list' condition the 'best' strategy was described as memorization by vocal rehearsal of a verbally presented list of color terms and spatial locations, see Figure 15, right. Second, a mapping acquisition phase using only unidimensional preceded blocks of conflict trials. During these mapping acquisition blocks, the verbal or spatial representation of the mapping rules was visible on the screen. Furthermore, throughout the rest of the experiment, whenever an erroneous response was given, the rule set was again shown on the screen in both instruction conditions.

Stimuli consisted of centrally presented colored words, taking up a visual angle of 1.66 to 2.22 degrees horizontally and 1.15 degrees vertically. Words were printed in a 24pt Helvetica font using uncapitalized letters. All stimuli were presented on a black background inside a fixation frame 3.01 degrees wide and 1.94 degrees high, whose four corners were visible all of the time. Depending on the task block, stimuli in the mapping acquisition trials consisted of the string XXXX printed in one of the relevant colors (in the Color task), or a color word displayed in white (in the Word task). In conflict trials, word color (red, blue, green, or yellow) could be congruent, neutral or incongruent to word meaning. The neutral condition in conflict trials used the same stimuli as the mapping acquisition trials.

Viewing distance was 70 cm. Responses were collected on the extended Macintosh keyboard using the keys 2, 4, 6, and 8 on the numeric keypad, which have a south/bottom, west/left, north/top, east/right layout.

The experiment was conducted in one session of approximately 45-60 min duration. There were two super-blocks of trials corresponding to the two tasks, Color and Word. Within each super-block, an initial mapping acquisition block was followed by a Stroop task-block of 360 trials, with an opportunity to rest after every 60 trials. New mapping acquisition trials were generated

---

[68] The low percentage of incongruent trials was due to a programming error that was detected too late. It has been shown that the size of the Stroop effect is larger when fewer incongruent trials are used, which is probably due to the fact that strategic reliance on the information provided by the distractor is advantageous when there is a relatively large proportion of congruent trials (see Logan, 1979). This strategic effect might be limited to cases of non-overlapping target-response ensembles. Because in the present experiment, targets had a color semantic, and responses were spatial, it seems reasonable to assume that the strategic influence is similar for responses based on word meaning and on word color.

until 24 consecutive correct responses had been given. During the first 24 mapping acquisition trials in each super-block, an equal proportion of the four color concepts was shown in random order. Additional trials were generated every four trials by randomly picking one of the four colors without replacement. During the mapping acquisition trials, mapping instructions were visible on the screen. Depending on instruction conditions either the verbal color-direction list was shown above fixation, or four colored squares were shown in locations around the stimulus that corresponded to the response keys, i.e. north, south, east, and west of fixation. Instructions were not visible during actual trials in test blocks, however, they were again shown after each erroneous response.

A trial started with the presentation of an empty fixation frame marked by its four corners, which served as a preparation cue. After a fixed cue-stimulus interval of 500 ms, the stimulus was displayed until a response was given. After a correct response, the screen was erased and the next trial started. After a wrong response, a reminder of the mapping rules which was identical to the one used during the learning phase was displayed together with the stimulus that elicited the wrong response. The mapping reminder display remained on the screen until the next trial started, but at least for 2.5 seconds. To request the next trial after an erroneous response and presentation of the mapping rule reminder, the correct answer had to be given.

Results

Reaction Times. After removal of errors and lag-1 errors, mean reaction times for each participant and design cell were calculated and outliers were removed by applying two criteria: reaction times shorter than 200 ms or longer than 4000 ms were discarded first, followed by reaction times that deviated for more than 3 SD from the cell mean per participant. Applying these criteria led to the removal of 2.1 % of the raw data (young: 1.9%, old: 2.3%). Means of the remaining reaction times per participant and design cell were entered into the analyses. The pattern of reaction times by condition is presented in Table 7.

Data were first analyzed globally, using a 2 (Age group) × 2 (Instruction) × 2 (Task) × 3 (Stroop condition) mixed mode analysis of variance with Task and Stroop condition as factors with repeated measures. In this analysis, it seems like the instruction manipulation was unsuccessful. Neither the main effect nor any interaction involving the instruction factor was significant. However, a closer inspection by taking task order into account revealed some very interesting interactions involving Age, Task and Instruction. Description of this second analysis will be postponed until the global results have been reported.

Significant main effects were obtained as expected for Age, $F(1,76) = 94.10$, $p < .001$, $MS_e = 70397.65$, Task, $F(1,76) = 38.06$, $p < .001$, $MS_e = 26935.92$, and for Stroop condition, $F(2,152) = 178.65$, $p < .001$, $MS_e = 15685.11$, with significant facilitation, $F(1,76) = 15.254$, $p < .001$, $MS_e = 1821.382$ and interference, $F(1,76) = 168.397$, $p < .001$, $MS_e = 46971.22$. Old adults were slower than young adults, reactions in the Word task were faster than in the Color task (mainly due to less interference, see below), and reactions on congruent trials were faster than on neutral trials, which were in turn much faster than reactions on incongruent trials.

There were several interactions involving Stroop condition. First, the interaction of Stroop condition with Age, $F(2,152) = 36.59$, $p < .001$, $MS_e = 15685.11$, which was limited to the interference contrast, $F(1,76) = 38.14$, $p < .001$, $MS_e = 46971.22$ (facilitation: n.s.), indicates that old adults were much more affected by incongruent stimuli than were young adults—a result entirely consistent with our earlier results suggesting larger age differences in the Stroop effect with an arbitrary mapping. Second, the interaction of Stroop condition and Task, $F(2,152) = 81.98$, $p < .001$, $MS_e = 5320.64$ in interference, $F(1,76) = 88.74$, $p < .001$, $MS_e = 15595.35$, but not facilitation (n.s.), indicates stronger interference by incongruent words in the Color task than by incongruent colors in the Word task—i.e. the Stroop effect was stronger than the reverse Stroop effect. However, the reverse Stroop effect was still substantial, at 93 ms for young adults, and 202 ms for old adults. In an analysis limited to the Word task, Stroop condition was therefore highly significant, $F(2,152) = 62.62$, $p < .001$, $MS_e = 8292.81$.

Third, and most interestingly, Age, Task, and Stroop condition interacted, $F(2,152) = 23.25$, $p < .001$, $MS_e = 5320.64$. The age difference in Stroop interference was larger in the Color task than in the Word task , $F(1,76) = 24.87$, $p < .001$, $MS_e = 387916.66$ (facilitation n.s.) The reported three-way interaction also qualifies the apparent interaction of Age and Task, $F(1,76) = 17.65$, $p < .001$, $MS_e = 26935.92$: age-related slowing was generally larger in the Color task than in the Word task, but particularly so in the incongruent condition of the Color task. These effects are interesting, because they support the view that slowing in the verbal domain is less severe than in the spatial domain. Therefore, a more detailed analysis looked at the Age×Task interaction in the neutral and congruent Stroop conditions, which was still significant, $F(1,76) = 6.64$, $p = .012$, $MS_e = 11471.31$. In these conditions, there was absolutely no Task effect for young adults (*Ms* =690 vs. 691 ms, Word vs. Color), $F < 1$, while old adults were significantly slower in the Color task (*Ms* =997 vs. 1060 ms), $F(1,38) = 8.75$, $p = .005$, $MS_e = 18183.45$. For old adults, retrieval of the matching spatial concepts thus appears to be faster if the stimulus uses a verbal code than if it uses a perceptual color code. Together with the large size of old adults' interference effects due to incongruent words in the Color task this suggests that they tend to use a phonological coding of the arbitrary mapping rules.

All other effects in this global analysis failed to reach significance.

Table 7
Means and standard errors [ms] for reaction time (columns 1-4), and mean error percentages (columns 5-6) in Experiment 6, broken up by Age, Instruction, Task, and Stroop condition.

			mean RT (s.e.) [ms]		*percent errors*	
Task	*Instruction*	*Stroop condition*	young	old	young	old
Color (Stroop)	spatial	congruent	670 (25.21)	1040 (47.68)	2.1	0.6
		neutral	690 (27.30)	1041 (44.06)	4.1	1.2
		incongruent	845 (40.90)	1489 (81.27)	6.1	9.0
	verbal	congruent	699 (23.99)	1079 (51.90)	2.4	0.7
		neutral	707 (25.41)	1081 (54.34)	3.7	1.3
		incongruent	873 (32.85)	1573 (91.26)	5.7	9.7
Word (reverse Stroop)	spatial	congruent	669 (26.95)	987 (54.49)	1.7	1.2
		neutral	703 (26.40)	995 (54.26)	2.1	0.8
		incongruent	774 (37.20)	1160 (66.27)	5.4	4.5
	verbal	congruent	690 (25.63)	991 (41.95)	2.1	1.0
		neutral	698 (23.76)	1016 (43.64)	2.9	1.0
		incongruent	772 (34.47)	1223 (81.83)	4.5	4.0

Several results of this analysis deserve discussion: First, the Instruction manipulation completely failed to influence responding. Thus it appears as if it was unsuccessful in establishing the intended differential verbal or spatial representations. Although a subsequent analysis including Task order (reported below) indicates that this might not be the full story, the weak effects of the instruction manipulation could indicate that participants tend to flexibly choose the code that they learn is most appropriate for a given task. In retrospect, it might not have been a clever aspect of the design to manipulate Task within subjects in a single session, because in theory, one and the same instruction was beneficial for the one task and detrimental for the other.

Second, substantial reverse Stroop effects were obtained in the Word task, although these were still much smaller than the Stroop effects in the Color task. While published results indicate that with a vocal response format, reverse Stroop effects are often absent, it appears that with the manual response format used here, irrelevant color distractors interfere with ‘naming’ of a word. Additionally, reverse Stroop effects were larger for old than for young adults, which indicates that the interference-proneness of old adults’ episodic retrieval processes is relatively independent of the type of distractor code. Responding in the Word task requires translation of a graphemical into a spatial code using an arbitrary mapping. The locus of interference depends on the working memory system used. With a verbal coding of rules, which appears to be more appropriate than spatial coding, an incongruent color can interfere during retrieval of the matching spatial concept, however, the phonological retrieval process will have a head-start, because its activation by the target word is fast, while interference by the color distractor in verbal working memory can only start after activation of the associated lexical entry. With a spatial coding of the rules, an incongruent color distractor would already interfere with the translation of the graphemical code into a conceptual color code. After translation, there will still be interference at retrieval. Generally, the relatively large reverse Stroop effects obtained in the current paradigm indicate that conceptual Stroop interference occurs during ‘episodic’ retrieval in a manual response task with arbitrary mappings.

The fact that Stroop interference is much larger than reverse Stroop interference could be interpreted to suggest that arbitrary rules tend to be coded in verbal working memory. However, the lack of an Instruction effect suggests that participants just use the code that is most appropriate for a given task, i.e., a spatial code in the Color task and a verbal code in the Word task. A more likely explanation for the interaction of Task and Stroop condition is therefore based on differences in the speed of conceptual activation by words and by colors, with words being faster.

Third, the Stroop effects and the Age difference therein were of a very large magnitude. The large Stroop effects might have been caused by the low percentage of incongruent trials. The large Age differences in the Stroop effect might also have been affected by this, however, a large effect had also been expected based on the fact that arbitrary mapping rules were used in all conditions of the present task.

Task Order Effects. The fact that the instruction manipulation apparently completely failed to influence responding was somewhat disappointing. However, an analysis taking task order into account hints at a possible problem old adults might have with proactive interference from earlier task sets. In fact, one possible interpretation of the results from the analysis to be reported below is that proactive interference completely reversed instruction effects in the second block task. Results are presented because they appear interesting, and suggest that instruction might have had an effect after all. Furthermore, a similar argument has recently been made in a comparison of high and low working memory span subjects (Kane & Engle, 2000). However, an advance note of caution appears appropriate: results from the task order analysis have to be interpreted very carefully, (a) because they were post-hoc motivated, and (b) because they are based on a between groups comparison. Task order is an additional between subjects factor, so that means are based on only 20 participants for each combination of Age group, Instruction condition, and Task Order.

Because of the between groups comparison, a first check was performed to make sure that the groups were comparable with respect to subject characteristics. In an analysis of variance using the digit symbol score as the dependent measure, and Age group, Instruction, and Task Order as between subjects factors, only Age was significant. Furthermore, if the analysis was limited to the old group, there were no significant effects of the factors Instruction and Task Order. Thus at least with respect to the digit symbol score, which has consistently been found to explain a large amount of age-related variance, the groups in the Order analysis were comparable. A further analysis was performed with the MWT-A score as the dependent measure, and again, except for the Age main effect, there were no further significant differences between groups. Nevertheless, to address the concern that the groups might be different, it was decided to include both the digit symbol and the MWT-A scores as covariates in the analysis. Of course, statistically controlling for digit symbol performance is expected to reduce the Age main effect.

Presentation of the results will proceed in two steps. First, results from a complete analysis incorporating all of the factors from the previous analysis plus Task Order are presented. Second, individual analyses of performance of first block and second block tasks are presented to evaluate effects of the instruction manipulation and of proactive interference due to first-block task sets, respectively.

The complete design was analyzed using a 2 (Task Order) × 2 (Age group) × 2 (Instruction) × 2 (Task) × 3 (Stroop condition) mixed mode analysis of covariance, with Task Order, Age group, and Instruction as between subjects factors, and Digit Symbol and MWT-A scores as covariates[69]. In the order analysis, there were several interactions involving Instruction and Task Order, which will be discussed in detail below. With respect to the effects not involving Task Order, most of the significant effects from the global analysis reported above were replicated. In particular, statistically controlling for digit-symbol score did not eliminate the Age main effect, $F(1,70) = 34.27$, $p < .001$, $MS_e = 52641.08$. Further effects that were similar in the detailed and in the global analysis include the main effect of Stroop condition, $F(2,140) = 5.36$, $p = .006$, $MS_e = 13821.08$ and the interaction of Age and Stroop condition, $F(2,140) = 21.87$, $p < .001$, $MS_e = 13821.08$.

Two effects that were significant in the global analysis failed to reach significance in the task order analysis, namely the main effect of Task, and the interaction of Task and Stroop condition. Thus in the task order analysis, the Stroop effect was not generally larger than the reverse Stroop effect. However, there were two higher-order interactions involving Task and Stroop condition. First, old adults did produce larger Stroop than reverse Stroop effects, as indicated by the significant interaction of Age, Task, and Stroop condition, $F(2,140) = 8.57$, $p < .001$, $MS_e = 4801.40$.

Second, there was an interaction of Task Order, Task, and Stroop condition, $F(2,140) = 8.23$, $p < .001$, $MS_e = 4801.40$, in the interference contrast, $F(1,70) = 7.64$, $p = .007$, $MS_e = 14320.34$ (facilitation n.s), which was independent of Age group. The reverse Stroop effect was larger if the Word task followed the Color task than if it came first. To a lesser extent, there was also a larger Stroop effect if the Color task followed the Word task than if it came first. Thus a history of reacting to one stimulus dimension seems to enhance its distracting quality once it becomes a distractor. This effect, which can best be interpreted as proactive interference, was observed in both age groups. The fact that the Order effect was stronger for the reverse Stroop task is consistent with results obtained by Allport et al. (1994; 2000), indicating that performance in an easy task is impaired after switching away from a difficult task. The alternative explanation, that a history of actively ignoring one dimension might reduce its quality as a target, appears less likely. Although an Order × Task interaction was obtained in an analysis limited to response times in the neutral conditions, $F(1,70) = 7.90$, $p = .006$, $MS_e = 6403.61$, this interaction is mainly caused by the fact that responding in the neutral condition Word task is actually faster if the task is performed second than if it is performed first. This seems difficult to reconcile with a hypothesis assuming that a history of actively ignoring the word causes it to be processed more slowly.

Let us now turn to the somewhat complex pattern of interactions involving Task Order and Instruction. Since Instruction had no effect in the global analysis, the effect of the manipulation was subtle—indeed it seems that its effects in the first block were completely counteracted by the second-block effects. The effect of Instruction is visible in a triple interaction of Task Order, Task, and Stroop condition, $F(2,140) = 4.53$, $p = .012$, $MS_e = 13821.08$, and in a four-way interaction of Task Order, Age, and Instruction with Stroop condition, $F(2,140) = 6.03$, $p = .003$, $MS_e = 13821.08$, in particular with the Stroop interference contrast, $F(1,70) = 6.33$, $p = .014$, $MS_e = 41986.03$; facilitation n.s.

Inspection of Table 8 shows that the four-way interaction is due to a disordinal interaction of Task Order and Instruction for old adults in the interference measure (incongruent-neutral), which is absent in young adults. This is graphically shown in Figure 16.

Let us call the degree to which instructions that are 'appropriate' for a given task according to the argument

---

[69] All of the reported effects were also significant if the covariates were not included.

Table 8
Experiment 6: Mean reaction times [ms] and error percentage by Age, Task Order, Task, Instruction, and Stroop condition. For Task Order, the label "Word, Color" indicates that the word (reverse Stroop) task was performed before the color (Stroop) task.

			mean RT [*ms*]				*percent errors*			
			young		old		young		old	
Task	*Instruction*	*Stroop condition*	Word, Color	Color, Word	Word, Color	Color, Word	Word, Color	Color, Word	Word, Color	Color, Word
Color (Stroop)	spatial	congruent	631	709	1109	971	1.9	2.3	0.6	0.6
		neutral	669	710	1096	985	4.2	3.9	1.0	1.5
		incongruent	810	880	1643	1336	6.8	5.4	12.4	5.7
	verbal	congruent	669	728	1054	1105	2.2	2.7	0.7	0.6
		neutral	692	723	1045	1116	3.7	3.6	0.7	1.9
		incongruent	895	850	1471	1676	6.9	4.4	6.7	12.6
Word (reverse Stroop)	spatial	congruent	659	679	1102	872	2.0	1.5	1.5	0.8
		neutral	700	705	1100	890	2.6	1.7	1.2	0.4
		incongruent	751	798	1282	1037	5.3	5.6	7.1	1.8
	verbal	congruent	698	682	1008	974	2.6	1.7	0.9	1.2
		neutral	707	689	1023	1008	3.3	2.4	0.6	1.4
		incongruent	751	793	1094	1352	2.8	6.3	0.8	7.1

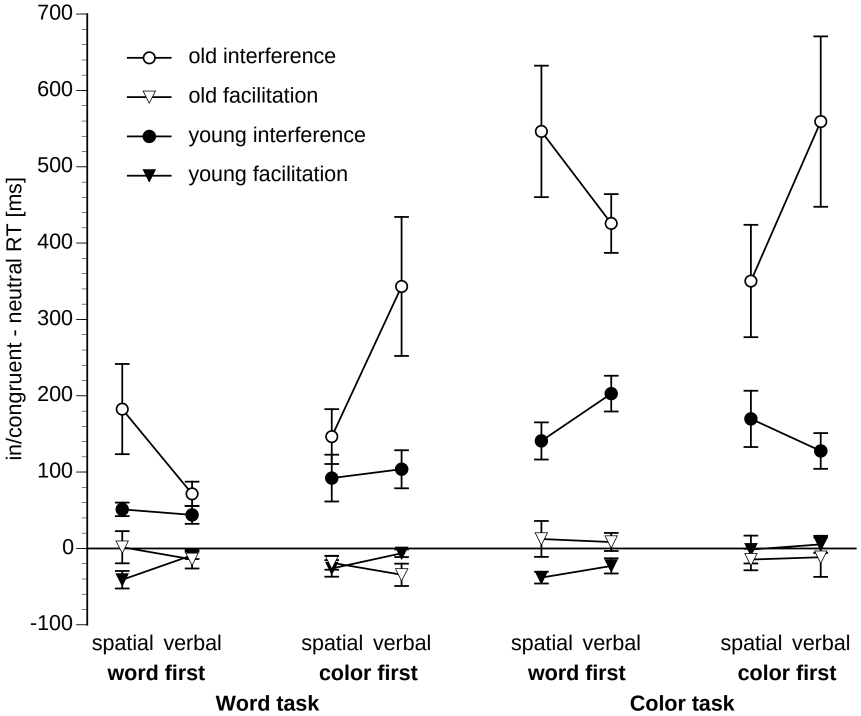


Figure 16. Experiment 6: Influence of Age, Instruction, Task, and Task Order on Stroop interference. One possible interpretation of this pattern is that old adults profit from instructions that suggest a task-appropriate coding for the arbitrary rules, but that this is limited to the task performed in the first block. In the first block, old adults' interference effects are smaller for the Color task if instructions suggest a maintenance in spatial working memory, and they are smaller in the Word task is instructions suggest a verbal coding of rules. In the task performed second, this pattern is reversed. One possible explanation is that the inappropriate instruction in the first task leads to a particularly strong task set for the task performed first, which is difficult to deactivate once the former target becomes a distractor. However, note that different groups of subjects were in the groups defined by task order and instruction. Thus an alternative interpretation is that the pattern is caused by sampling error. The old adults in the verbal-instruction group that started with the color task, and the old adults in the spatial-instruction group that started with the word task might have been particularly interference-prone.

brought forward in the introduction to this experiment task-instruction compatibility. Task-instruction compatibility is relatively high for a spatial instruction in the Color task, and for a verbal instruction in the Word task, compared to the other two combinations. It appears as if old adults performed better and in particular suffered less from interference with task-compatible instructions, but only if the task was performed first. If the task was performed after the other (instruction-incompatible) task, then old adults actually performed worse than the group that started with the incompatible instructions.

First block, second block. To facilitate description of the four-way interaction, results from the first task block and from the second task block were subjected to separate Age×Instruction×Task×Stroop condition ANOVAs, taking Task as a between-subjects factor, and again controlling for digit symbol and MWT-A scores. With regard to the effect of task-instruction compatibility, an interaction of Task and instruction was expected. For the task performed in the first block, the instruction manipulation had the expected effect, albeit it was limited to old adults: the spatial instruction helped responding in the Color task, and the verbal instruction helped responding in the Word task (see Figure 17). Although the pattern of means suggests that both the match of target dimension code and working memory code, and the match of distractor code and working memory code might have contributed to this effect, the first point cannot firmly be made. Numerically, in the Color task, the spatially instructed group of old adults was faster than the verbally instructed group (*M*s=1097 vs. 1299, respectively). In the word task, the reverse was true (*M*s=1161 vs. 1042, respectively). However, the interaction of Task and Instruction was only marginal, $F(1,70) = 3.27$, $p = .075$, $MS_e = 29695.50$, and the triple interaction with Age failed to reach significance, $F(1,70) = 1.30$, $p = .258$, $MS_e = 29695.50$. Furthermore, if only old adults' reaction times in the *neutral* Stroop condition are analyzed, the interaction is far from significant, $F < 1$,

Thus it appears that the main cause for the instruction effect are differences in distractor strength that are elicited by the instruction manipulation, chiefly for old adults. Interference is stronger if the distractor matches the code that is used for maintenance of the mapping rules. In the tasks performed in the first block, Stroop effects were smaller with task-compatible than with task-incompatible instructions, as indicated by the significant interaction of Instruction, Task, and Stroop condition, $F(2,140) = 3.37$, $p = .038$, $MS_e = 9827.19$. This interaction is further modulated by Age: old adults experienced significantly less interference with task-compatible than with incompatible instructions. This is indicated by the significant Age × Instruction × Task × Stroop condition interaction, $F(2,140) = 3.42$, $p = .035$, $MS_e = 9827.19$. In the Color task, spatially instructed old adults were less disturbed by incongruent word distractors than the verbally instructed group, while in the Word task, spatially instructed old adults were more disturbed by incongruent color distractors than the verbally instructed group. Instruction had no significant effect in an analysis limited to young adults. In contrast, if the analysis of tasks performed first was limited to old adults, the interaction of Instruction, Task, and Stroop condition was significant, $F(2,68) = 3.55$, $p = .034$, $MS_e = 16983.88$.

The spatially instructed group of old adults in the Word task was more influenced by incongruent color distractors than the verbally instructed group, while the verbally instructed group in the Color task suffered more from word distractors than the spatially instructed group. Thus for old adults, Stroop effects were smaller when instructions suggested using spatial codes for the target-response mapping in the Color task, and using verbal codes in the Word task, than with the reverse combinations. This supports the idea that the differential use of working memory codes suggested by the instructions mainly enhanced the strength of a distractor. If the instruction suggested a code that matched the distractor better than the target, old adults were particularly susceptible to Stroop interference. It appears as if interference in working memory is a major locus of Stroop interference for old adults in tasks that require an arbitrary mapping of stimuli to responses.

Now consider the tasks performed in the second half of the experiment. Again, the interactions of Instruction, Task, and Stroop condition, $F(2,140) = 3.36$, $p = .038$, $MS_e = 8905.24$, and of the three factors and Age, $F(2,140) = 5.77$, $p = .004$, $MS_e = 8905.24$, were significant. Again, effects of the instruction manipulation were mainly restricted to old adults. For them, the pattern completely reversed in comparison to the first-block task: interference in the tasks performed second was larger with task-compatible than with task-incompatible instructions. For example, the spatial-instruction group experienced (slightly) more interference in the Color task than the verbal-instruction group if the Color task was performed after the Word task. The verbal-instruction group experienced pronouncedly more interference in the Word task than the spatial-instruction group if the Word task was performed after the Color task.

At least two explanations for the fact that the pattern is entirely reversed for performance in the second-block task are conceivable. One possibility is that the effect is simply due to differences between the groups of participants. In particular, two groups of old adults might have consisted of individuals that were particularly susceptible to interference, namely the verbal-instruction group that started with the Color task, and the spatial-instruction group that started with the Word task. Although Instruction and Task Order do not interact in an analysis of old adults' digit symbol scores, this sampling-based explanation cannot be ruled out.

However, an alternative explanation is based on the hypothesis that under certain circumstances, proactive interference is particularly strong. This could be the case if old adults are confronted first with a task that is incompatible to the instruction. If for example verbally instructed subjects start with the Color task, this

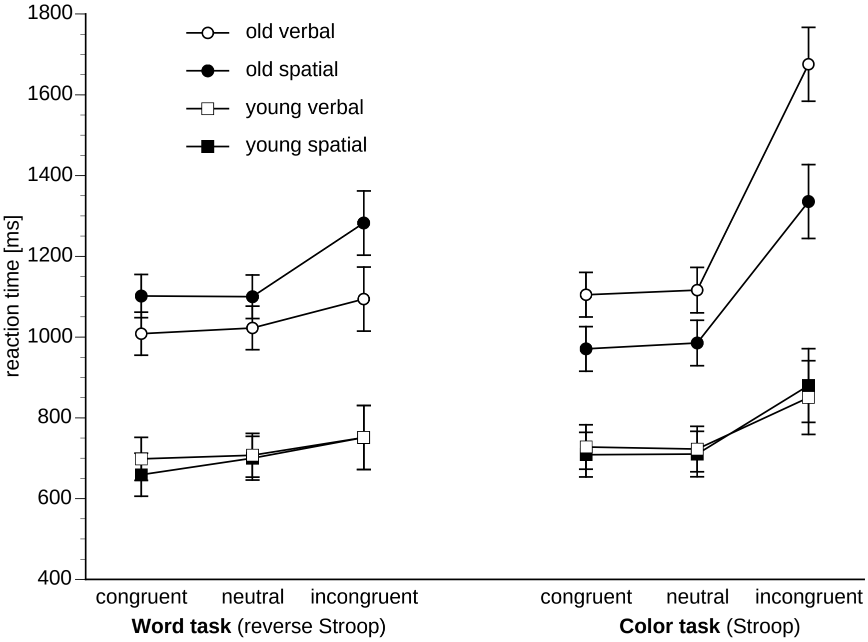


Figure 17. Experiment 6: Reaction times for the tasks performed first, broken up by Age, Instruction, Task, and Stroop condition. In the Color task, older adults appear to be faster and less susceptible to interference with a spatial than with a verbal instruction. In the Word task, instruction has the reverse effect. Assuming that the effect is due to the instruction manipulation and not due to group differences, old adults perform better if they use a memory code that matches the target code, and if they avoid a memory code that matches the distractor code.

may lead to the build-up of a particularly strong task set, because good performance in the task requires inhibition of the distracting task that the instruction is more compatible with. If more executive resources have to be spent in the first task to maintain the task-set, then this might lead to a task set that is more difficult to overcome once target and distractor have switched their role. The fact that proactive interference does play a role in the current paradigm is clearly visible in the interaction of Task, Task Order, and Stroop condition, that was obtained independent of Age. Stroop effects caused by an incongruent distractor are larger when the distractor dimension had formerly been the target. Now imagine old adults starting with an instruction that makes coding of the target-response rules difficult, because the distractor is more compatible with the memory code used to perform the first task. To overcome the potentially strong interference in the first task, these subjects might recruit a large amount of 'executive resources' or 'mental energy', which might in turn lead to a rather strong task set. In other words, the association of target-response rules might be particularly strongly associated with a given *task*, i.e. a value of the target dimension. This strong task set might then be difficult to de-activate in the task performed second. Hence in the second task, the relatively strong associations of task and mapping rule set acquired in the first task might continue to influence responding. This will impair performance in the second task, because the former target now acts as the distractor.

Proportional Measures. All effects involving Age that were significant in the global analysis remained significant in the analysis of log reaction times. These effects include the Age main effect, $F(1,76) = 126.54$, $p < .001$, $MS_e = 5.56\text{e-}2$, and the interactions of Age and Task, $F(1,76) = 14.16$, $p < .001$, $MS_e = 1.95\text{e-}2$, of Age and Stroop condition, $F(2,152) = 23.04$, $p < .001$, $MS_e = 6.54\text{e-}3$, and the triple interaction of Age, Task, and Stroop condition, $F(2,152) = 13.09$, $p < .001$, $MS_e = 3.03\text{e-}3$.

Similarly, if Task Order is taken into account, the same pattern of interaction as in the analysis of untransformed reaction times is obtained in the analysis of log reaction times (again including digit symbol and MWT-A scores as covariates). Significant effects involving Age include the Age main effect, $F(1,70) = 49.22$, $p < .001$, $MS_e = 4.12\text{e-}2$, interactions of Task and Age, $F(1,70) = 4.04$, $p = .048$, $MS_e = 1.95\text{e-}2$, Age and Stroop condition, $F(2,140) = 14.96$, $p < .001$, $MS_e = 6.08\text{e-}3$, and of Age, Task, and Stroop condition, $F(2,140) = 5.80$, $p = .004$, $MS_e = 2.61\text{e-}3$. Most importantly, the four-way interaction of Age, Instruction, Task Order, and Stroop condition remained significant in the log RT analysis, $F(2,140) = 4.41$, $p = .014$, $MS_e = 6.08\text{e-}3$.

The second proportional analysis used relative interference scored, calculated for each subject as the difference between incongruent and neutral condition reaction time in a given task, divided by the neutral condition reaction time.[70] Again, results from the analy-

---

[70] Note that an interference difference measure is analyzed, therefore a significant main effect indicates an interaction with Stroop interference in the untransformed measure (for example, the Age main effect corresponds to an interaction of Age

sis of proportional interference scores support the untransformed reaction time analysis. Significant age-related effects include the Age main effect, $F(1,76) = 20.97$, $p < .001$, $MS_e = 4.28\text{e-}2$, and the interaction of Age and Task, $F(1,76) = 10.76$, $p = .002$, $MS_e = 1.71\text{e-}2$. Thus over-proportional age effects in Stroop interference were obtained (a) in general[71], and (b) in the Color task in particular.

In the analysis of Task Order effects including the DSS and MWT-A covariates, the discussed effects involving Stroop condition all remained significant with the proportional interference measure. This is indicated by the main effect of Age, $F(1,70) = 14.12$, $p < .001$, $MS_e = 4.12\text{e-}2$, and the interactions of Task and Age, $F(1,70) = 5.93$, $p = .017$, $MS_e = 1.53\text{e-}2$, and of Age, Instruction, and Task Order, $F(1,70) = 4.64$, $p = .035$, $MS_e = 4.12\text{e-}2$. As an aside, the 'proactive interference' interaction of Task, Task Order, and Stroop condition was also significant in the analysis of proportional interference scores, as the interaction of Task and Task Order indicates, $F(1,70) = 11.35$, $p = .001$, $MS_e = 1.53\text{e-}2$.

Thus all the conclusions from the untransformed reaction times analysis can be transferred into proportional measurement space.

Errors. The overall error rate was low, with an average of 3.2%. For a classification of error rates by experimental condition see Table 7, and for a classification broken up by Task Order consult Table 8. The pattern of results in error analyses largely followed the reaction time pattern, although some effects were not significant in errors, namely the Age main effect and the interaction of Age and Task.

The main effects of Task, $F(1,76) = 19.18$, $p < .001$, $MS_e = 1.02\text{e-}3$, and of Stroop condition, $F(2,152) = 59.30$, $p < .001$, $MS_e = 1.69\text{e-}3$ were significant. More errors were committed in the Color task than in the Word task, and Stroop interference, $F(1,76) = 51.91$, $p < .001$, $MS_e = 4.89\text{e-}3$, and facilitation, $F(1,76) = 22.44$, $p < .001$, $MS_e = 2.95\text{e-}4$, were observed.

Age and Stroop condition interacted, $F(2,152) = 7.51$, $p = .001$, $MS_e = 1.69\text{e-}3$. Facilitation was larger for young adults, $F(1,72) = 10.25$, $p = .002$, $MS_e = 2.95\text{e-}4$, while interference (and the overall Stroop effect) was larger for old adults, $F(1,72) = 9.63$, $p = .003$, $MS_e = 4.89\text{e-}3$. In fact, without the high error rate on incongruent conditions, a significant age main effect would have been obtained, $F(1,76) = 37.48$, $p < .001$, $MS_e = 5.87\text{e-}4$, indicating that old adults produce less errors than young adults in the congruent and neutral conditions (like in previous experiments using manual responding).

Finally, there were interactions of Task and Stroop condition, $F(2,152) = 9.94$, $p < .001$, $MS_e = 1.03\text{e-}3$, and of Age, Task, and Stroop condition, $F(2,152) = 8.51$, $p < .001$, $MS_e = S1.03\text{e-}3$. Like in reaction times, the Stroop effect in errors was larger than the reverse Stroop effect. Furthermore, old adults produced a particularly large number of errors (9.3%) in incongruent conditions of the Color task, while their reverse Stroop effect was much smaller, with an error rate of 4.2% in the incongruent Word task. For young adults, there was a difference of less than one percentage point between the error rates in the incongruent Color and Word tasks (*Ms* =5.9% vs. 5.0%, respectively).

Task Order Effects. In the analysis including Task order effects, the pattern of results in error analyses also largely followed the reaction time pattern. In particular, the direction of the four-way interaction of Task Order, Age, and Instruction with Stroop condition in the interference contrast was the same, $F(1,72) = 5.21$, $p = .025$, $MS_e = 4.61\text{e-}3$. Other effects that were similar to the reaction time effects included the main effect of Stroop condition, $F(2,144) = 62.39$, $p < .001$, $MS_e = 1.59\text{e-}3$ in both facilitation, $F(1,72) = 20.67$, $p < .001$, $MS_e = 3.18\text{e-}4$, and interference, $F(1,72) = 54.24$, $p < .001$, $MS_e = 4.61\text{e-}3$, the main effect of Task, $F(1,72) = 18.15$, $p < .001$, $MS_e = 3.40\text{e-}4$, and the interactions of Age and Stroop condition, $F(2,144) = 7.90$, $p = .001$, $MS_e = 1.59\text{e-}3$, of Task and Stroop condition, $F(2,144) = 9.77$, $p < .001$, $MS_e = 9.71\text{e-}4$, of Age group, Task, and Stroop interference, $F(1,72) = 8.86$, $p = .004$, $MS_e = 2.84\text{e-}3$ (facilitation: n.s.), and of Task Order, Instruction, and Stroop interference, $F(1,72) = 6.61$, $p = .012$, $MS_e = 4.605\text{e-}3$ (facilitation: n.s.).

However, there were also certain differences in the pattern of significance. In the error analysis, old and young adults did not differ globally (Age main effect: n.s.). Neither was there an interaction of Age and Task. On the other hand, a new interaction of Instruction and Task Order, $F(1,72) = 7.31$, $p = .009$, $MS_e = 7.66\text{e-}4$, emerged in the analysis of errors that was not visible in the analysis of reaction times. This disordinal interaction indicates that overall, less errors were committed with a spatial instruction if the Task Order was Color first, Word second than with the reverse Task Order, while the opposite pattern was obtained with verbal instructions, which elicited less errors in the Word-Color than in the Color-Word task order. Furthermore, the interaction of Age and Stroop condition was no longer limited to interference, $F(1,72) = 10.14$, $p = .002$, $MS_e = 4.61\text{e-}3$, but now also included facilitation, $F(1,72) = 10.32$, $p = .002$, $MS_e = 3.18\text{e-}4$. For old adults, the difference in error rate between neutral and incongruent trials was larger than for young adults, while the difference between congruent and neutral trials was larger for young than for old adults, which were already close to the floor on neutral trials. It should also be noted that during congruent and neutral trials, old adults committed hardly any errors (about 0.6%), while young adults committed some (about 2.6%). On incon-

---

and Stroop interference). Division of the difference by the neutral condition reaction time accounts for interindividual differences in baseline speed.

[71] If the analysis was limited to the Word task, the Age main effect was still significant, $F(1,76) = 6.78$, $p = .011$, $MS_e = 1.98\text{e-}2$

gruent trials, however, old adults committed more errors than young adults (6.7% vs. 5.4% , respectively).

Discussion

An important result obtained in the present experiment was the large size of the reverse Stroop effect with manual responses. Compared to the typically negligible reverse Stroop effect obtained with vocal responding, this indicates that interference at episodic retrieval substantially contributes to Stroop-like interference, at least when a direct route from the relevant stimulus dimension to the response is not available.

A second important result was the observation that age effects in Stroop and reverse Stroop interference were large and could not be explained by differences in baseline response speed. In the light of results from the compatible conditions of Experiments 2–5, where interactions of age and Stroop condition were small or absent, this is likely attributable to the arbitrary S–R mapping employed in the present experiment and therefore adds to the evidence for a specific age-related deficit in episodic memory.

Unfortunately, results were less clear with respect to the main experimental question, whether old adults tend to rely on verbal short-term memory. The instruction manipulation completely failed to influence responding in the global analysis. One reason for this could be that subjects tend to flexibly choose a strategy to use the internal code that best suits a given task. In the current experiment, except for the instruction manipulation, no means were taken to ensure that subjects did not switch from the instructed code to a different code once the Stroop task had started. Furthermore, they even had the chance to redundantly code the mapping rules, using both verbal and spatial working memory. In working memory experiments, the spatial or verbal compartments are often selectively disabled, for example by using a concurrent memory load technique, or by having subjects continuously articulate nonsense syllables, thereby eliminating subvocal rehearsal. In the current experiment, no such techniques were used, because it was felt that they could increase the executive, coordinative demands of the task.

On the other hand, the task order analysis suggests a possible influence of instruction on old adults' pattern of responding. Briefly, in the task performed first, the groups of old adults that started with instruction-compatible tasks (Word task with verbal instruction, Color task with spatial instruction) were less susceptible to Stroop interference than groups of old adults that started with instruction-incompatible tasks. However, the pattern was reversed for the tasks performed second: if an instruction-compatible task was performed after an instruction-incompatible task, then interference effects for old adults were actually higher than for the instruction-incompatible task performed after the instruction-compatible task. Two alternative explanations are suggested for the pattern of order effects caused by the instruction manipulation. First, it might be that it was simply caused by individual differences in the groups of old adults that started with instruction-compatible and with instruction-incompatible tasks. Specifically, the verbally instructed group of old adults who started with the Color task, and the spatially instructed group that started with the Word task might have consisted of individuals with a particularly high susceptibility to interference. Although there were no group differences with respect to age, digit-symbol, or vocabulary score between old adults starting with task-compatible and task-incompatible instructions, the results are nevertheless based on a between subjects comparison, hence an explanation based on individual differences cannot be ruled out.

In retrospect, the design choice to manipulate Task within subjects in a single session was unfortunate. Why this is so becomes obvious if the second explanation for the interaction of instruction effects and task order is considered, which is based on a combination of two different, but related mechanisms. The first mechanism is relevant for the effects obtained in the task performed first. It is based on the relative match of target and distractor codes to internal, working memory codes and is therefore related to task-instruction compatibility. Recall that Durgin (2000) found a complete reversal of the Stroop and reverse Stroop effect in a color-matching task. Based on these results, I expected to find a similar effect with regard to the instruction manipulation. With a spatial instruction, the typically large difference in magnitude between Stroop and reverse Stroop effects was expected to be reduced. This pattern was in fact obtained, but limited to the tasks performed first and to the old group. In the task performed first, old adults experienced less interference from incongruent distractors if the instruction suggested the use of an internal code that matched the target dimension than if the code matched the distractor dimension. That is, color distractors caused more Stroop interference with a spatial than with a verbal instruction, and word distractors caused more reverse Stroop interference with a verbal than with a spatial instruction.

The second mechanism that might be responsible for the interaction of instruction and task order had not been anticipated when the experiment was designed. It is related to proactive interference and will be discussed in somewhat greater detail below, because it might be able to explain the paradox task-instruction compatibility effects produced by old adults in the tasks performed second. The fact that proactive interference does play a role in the current paradigm, independent of possible group differences, can be illustrated by the result that incongruent distractors cause larger interference effects if they have been targets before. Stroop interference effects were larger if the Color task was performed after the Word task than if it was performed first, and similarly reverse Stroop effects were larger if the Word task was performed after the Color task. However, this kind of strengthening of the distractor by its history as a target was independent of age group and instruction condition. Having established that proactive interference is potentially important in the current paradigm, let us now turn towards another effect that can be interpreted as proactive interference, and that might be sensible to the instruction manipulation. This effect, which is limited to the old group, is indexed by the fact that in the group of old adults, for tasks performed second, interference in instruction-compatible tasks was larger than in instruction-incompatible tasks. How can this pattern of paradox task-instruction compatibility effects be explained by proactive interference? The explanation is based on an argument developed by Kane, Bleckley, Conway, and Engle (2001) (see also Kane & Engle, 2000), which realizes that the amount of proactive interference might be positively correlated with the strength of a task set. Kane et al. (2001) tested individuals differing in working memory (WM) span on both pro- and antisaccade tasks. Location of a target was cued by a flashing symbol either at (prosaccade task) or opposite (antisaccade task) the target position. Kane et al. found no group differences in prosaccade performance when it was the first task tested, which they interpret as equivalence in automatic orienting. However, the low-span group was significantly slower and less accurate in the antisaccade task, especially if it was the first task tested. Kane et al. assume that this is due to differences in attentional control correlating with differences in WM span. The most interesting result for the current investigation is that performing the antisaccade task first impaired performance in the subsequent prosaccade task, and especially so for low-span individuals. This can best be explained by the concept of proactive task-set interference. The 'automatic' prosaccade task, which is not normally affected by a concurrent memory load and may be performed with little involvement of controlled attention, may be disrupted by the prior performance of a similar, but attention-demanding task. For low-span individuals, switching intentional set from the controlled antisaccade task to the prosaccade task appears to be quite difficult.

The results resemble those of Allport and colleagues (1994; 2000), who have argued that proactive task-sets interference is a major source of switch costs observed in the task-switching paradigm, which is regarded as one of the prime paradigms for tapping executive processes. Allport et al. (1994) examined task switching in a series of experiments using various Stroop-like tasks. In their Experiment 5, switch costs turned out to be asymmetrically higher when switching away from a hard, 'controlled' task (Stroop color naming) towards an easy, 'automatic' task (Stroop word reading), than in the reverse case. Whether the task set inertia variant of proactive interference is visible in both automatic and controlled tasks is a topic of current debate (Rogers & Monsell, 1995; Allport et al., 1994; Allport & Wylie, 2000; Kane et al., 2001). Some authors argue that a prerequisite for task set inertia is a switch away from a task requiring controlled attention. For example, Wylie and Allport (2000) showed that proactive task-set interference depends on the amount of control required by the task that is switched away from.

In summary, there are several observations reported in the literature suggesting that switching from a more automatic to a more controlled task causes minimal difficulty compared with switching from a controlled to an automatic task. This effect is apparently exacerbated in individuals with working memory deficits. In general, a task that is more difficult in the sense that it requires a higher degree of executive control seems to lead to a stronger task set. Once a strong task-set has been established, it can interfere with a weaker task-set in a different task that is subsequently performed (or alternated) within the same stimulus-response context. [72]

If the strength of a task set tends to persistently influence responding even after a task has become irrelevant, then this might be able to explain the paradox task-instruction compatibility effects produced by old adults in the task performed second. The asymmetry of proactive interference should be measurable in 'global switch costs' (Kray & Lindenberger, 2000; Mayr, 2001; Mayr & Liebscher, 2001; Meiran et al., 2001), leading to the relative impairment of a 'weaker-set' task if it follows a 'stronger-set' task.

Note that only two orders of task-instruction compatibility were realized: either an instruction-compatible

---

[72] Allport et al. (1994) discuss this effect under the label of task set inertia, a kind of proactive interference, in which a nondominant response mapping imposes a stronger set that is more difficult to overcome or deactivate than is the set for the dominant response. In the current experiment, dominance of the S-R mapping was held constant: only one response mode was used, with no pre-experimental associations to any of the stimuli. In other words, pre-experimentally established differential automaticity does not play a role in the current setting, unlike in the pro-/antisaccade task. Since the stimulus-response mapping was arbitrary for both tasks, we presume that none of the tasks is highly automatic, and thus task set inertia is expected to be observed in switching away from either. If differential automaticity did not contribute, what else could have induced task-sets of different strengths? It is conceivable that in the current experiment, task-sets of different strengths were induced by instructions differing in degree of compatibility with the task to be performed first.

Based on the result that a reverse Stroop effect can be obtained in a color-matching Stroop task (Durgin, 2000), it was assumed that a task-set compatible to the color task could be induced by presenting a spatial layout of color patches, while a task-set more compatible to the word task could be induced by the verbal list instruction. An instruction that was less well suited to the task at hand might then have led to the establishment of a stronger task-set. A strong task-set is likely to be built up and maintained during an instruction-incompatible first block not only because of the translation required to map stimuli onto the input side of the mapping rule, but also because distractors are more compatible with the code used to maintain the mapping rules than targets. Protection from interference from instruction-compatible distractors might be critical for establishment of a strong task-set.

Establishment and maintenance of task-set might require a smaller amount of executive control with a task-compatible instruction, while having to establish and maintain a relatively task-incompatible set of mapping rules in the first task should require more executive resources. Thus depending on the match between instructions and task, either a weak or a strong task-set might have been built up during the first task block.

task followed an instruction-incompatible task, or an instruction-incompatible task followed an instruction-compatible task. In the present study, proactive task-set interference would therefore be indicated by higher reaction times and higher interference effects in the instruction-compatible second-block task (that is performed after the instruction-incompatible task), (a) if compared to the situation where it is performed as the first task, and (b) compared to the instruction-incompatible second-block task (that is performed after the instruction-compatible task). With spatial, color-matching instructions, the Color task should be impaired if it follows the Word task, because the instruction did not match the task performed first and hence caused a relatively strong task-set to be built up in the first block. Similarly, with word-reading instructions, the word task should be impaired if it follows the color task. The effect should be particularly pronounced under conditions of Stroop interference, because interference resolution is typically considered to be tapping executive resources.

Encountering an instruction-compatible task in the second block means that it was performed after a strong task-set has been built up for the instruction-incompatible first block task. This task-set might interfere with the second block task, particularly because the stimuli did not change between blocks (except in the neutral Stroop condition). There is no principle reason why this kind of proactive task-set interference should be limited to old adults. To the degree that the task is taxing enough to limit availability of prefrontal, executive resources even for young adults, they are expected to show a qualitatively similar pattern. However, it has previously been shown that young adults are far more flexible that old adults when it comes to adapting to new tasks (e.g., Mayr & Liebscher, 2001). The fact that young adults did not show an instruction effect in the task performed first might be indicative of this flexibility.[73] Furthermore, larger proactive interference effects for old adults were to be expected based on results showing that old adults are more susceptible to intrusion errors (e.g., Oberauer, 2001) and produce larger global switch costs (Kray & Lindenberger, 2000; Mayr, 2001; Mayr & Liebscher, 2001; Meiran et al., 2001) than young adults. Finally, according to Kane et al. (2001) proactive task-set interference effects might be limited to or at least more persistent in individuals with a low working-memory span, in our case old adults.

It seems likely that proactive interference due to strong task sets involves similar systems as the establishment and maintenance of an arbitrary S-R mapping. After all, a task set is by definition a set of rules required to perform a task at hand. Thus mapping rules are part of a task set, which in addition to mapping rules encompasses further task-relevant aspects, such as goals and intentions of the actor. Here, we focus on the latter components. A strong task set, established for example by a task requirement to overcome a habitual response tendency as in the antisaccade task, or by a history of resolving relatively strong interference caused by a mental coding of mapping rules in a format that matches the distractor better than the target, impairs performance in a subsequent task using the same stimuli, but better-suited task rules. Conversely, task instructions that are more compatible to the task at hand lead to the establishment of a weaker task set, with a negligible influence on subsequent performance of a 'strong-set' task.

To conclude, an admittedly post-hoc explanation was proposed to account for the interaction of Instruction, Task Order, and Stroop condition observed in the group of old adults. The explanation based on the match of task codes and internal codes can account for the fact that old adults show task-instruction compatibility effects in the task performed first. If only the task performed first was analyzed, old adults were sensitive to the instruction manipulation. They performed better in tasks for which the target code matched the instructed working memory code than in tasks in which the distractor code matched the working memory code. If a strong task set was induced in the task performed first, by instructing a working memory representation that matches the distractor code, then this led to a decrease in performance in the task performed in the second half for old adults, who have repeatedly reported to be highly susceptible to proactive interference (or intrusion errors, e.g. Oberauer, 2001), and who are much more likely to suffer from task set inertia (e.g., Mayr & Liebscher, 2001; Meiran et al., 2001). With the additional assumptions that (a) the strength of the task-set built up during the first block is negatively correlated with the degree of compatibility of (first-block) task and instruction, (b) a strong task-set proactively interferes with a subsequent task that is performed with the same set of stimuli and responses, and (c) old adults are more susceptible to proactive interference than young adults, possibly due to their reduced working memory capacity, the explanation can also account for the paradox task-instruction compatibility effect obtained in the older groups in the task performed second.

However, let me again clearly state that the explanation was only brought up post-hoc. Furthermore, because it is based on results of a between-groups comparison, a more parsimonious alternative explanation based on individual differences is available. Clearly, additional research is needed to support both the code compatibility and especially the proactive task-set interference hypotheses. For a start, it seems desirable to replicate the experiment, but manipulate task between sessions, so that the first task-set can decay between sessions.

General Discussion

Synopsis

In this research, it was proposed that a model assuming specific effects at rather high-level processes related to episodic and working memory can better explain age differences in cognition than a general, unspecific slowing model. Specifically, the hypothesis was brought for-

---

[73] Alternatively, it might just indicate that they failed to follow instructions.

ward that the age-by-complexity effect, which is indicated by a regression slope larger than one in the Brinley plot, might be explained by a covariation of memory demands and task complexity. It was argued that the contribution of executive, working memory or episodic memory processes to reaction time is often neglected, but that it might be relevant in all but the most trivial tasks and might furthermore be one of the major sources of between task-variance in young adults reaction time.

The Episodic Accumulator Model was proposed as an alternative to models of age-related slowing that assume a single, low-level deficit. The Episodic Accumulator Model assumes a specific age deficit in memory processes that are involved in the retrieval and maintenance of freshly acquired associations tied to a particular context. More generally, it assumes that age differences vary with the degree of arbitrariness in the set of task rules.

To test the model in a series of experiments, it was necessary to independently vary 'late' memory demands and a separate, 'early' factor affecting cognitive difficulty. The reasoning was that according to the model, manipulation of the memory factor would produce large age-by-condition interactions and steep Brinley slopes, while manipulation of the early difficulty factor at the level of low memory demands would produce small age-by-condition effects and Brinley slopes close to one. The influence of the early factor at the level of high memory demands was predicted to depend on the degree to which 'early' and 'late' processing overlap in time. In the case of serial, stage-like processing, early and late difficulty effects were predicted to be independent with regard to age effects. In the case of parallel, cascaded processing, when the effects of the early difficulty manipulation cannot fully be resolved before memory retrieval starts, effects caused by the early difficulty manipulation were predicted to be age-differentially amplified by high memory demands.

Experiment 1 varied perceptual difficulty on the one hand and arbitrariness of the S-R mapping, i.e., episodic memory demands, on the other hand. The memory factor caused larger age effect than the perceptual factor. Both age groups were slower with an arbitrary than with a compatible mapping, and the increase in the age effect with the arbitrary mapping was over-proportional. The manipulations of perceptual difficulty and of memory demands were perfectly additive with respect to age effects, which indicates that there is not much temporal overlap between perceptual classification and episodic retrieval. Furthermore, results can be interpreted to indicate that the majority of age effects arise only after perceptual classification, which is consistent with independent results from ERP research. Experiment 1 also included an inverse mapping condition, using an incompatible S-R mapping that posed relatively low episodic demands. The inverse mapping led to identical age effects as the compatible mapping, a result that points to episodic demands, not response inhibition, as a source of age-related slowing.

In Experiments 2-6, Stroop condition was manipulated as an early difficulty factor that was assumed to produce more temporal overlap with memory retrieval. In Experiment 2 vocal or manual responses were given to the color of Stroop color-word stimuli. Here, episodic demands were manipulated by varying the response modality: they were larger with manual responses, which in contrast to vocal color responses are not pre-experimentally associated with color concepts. Results generally supported the Episodic Accumulator Model predictions. Slowing was over-proportional with, and Stroop effects were age-differentially amplified by the arbitrary mapping: Young adults experienced far less Stroop interference with manual than with vocal responses, while for old adults, the amount of interference did not differ between response modalities. Additionally, results from a comparison of tasks featuring integrated or separated color-word ensembles indicate that the age-differential amplification of early difficulty effects by memory demands is larger when early attentional filtering is not possible.

Experiment 2 set the stage for a number of follow-up experiments. Experiments 3 and 4 addressed the confounding of memory demands and response modality inherent in the mapping manipulation chosen in Experiment 2. Experiment 3 compared compatible and arbitrary mappings using vocal responding, with the arbitrary response set consisting of male first names. Results were not entirely decisive. At the global level, age differences were not affected by memory demands, i.e. Age and Mapping did not interact. Even in the absence of this interaction, a numerically larger age difference in the Stroop effect was obtained with the arbitrary than with the compatible mapping, however, this tendency was not significant. A more detailed look showed that strong repetition effects modulated the pattern. In particular, repetition effects also modulated the age difference in the Stroop effect. A target repetition in the arbitrary mapping condition means that the rule that was applied on the previous trial is still active, hence does not have to be retrieved on the current trial. In fact, responses on repetition trials were as fast with the arbitrary as with the compatible mapping. If retrieval of arbitrary rules is a critical moderator of age differences, then predictions of the Episodic Accumulator Model might be limited to the change trials. Indeed, the age difference in the Stroop effect was independent of mapping condition on repetition trials, whereas it was significantly more pronounced in the arbitrary than in the compatible mapping condition on change trials. Experiment 4 replicated the comparison of compatible and arbitrary mappings in the vocal modality, using a more restricted arbitrary response set of the natural numbers one to four. Experiment 4 also reintroduced a manual response condition, thereby replicating Experiment 2. Together, this allowed to compare (a) the effect of episodic memory demands (compatible vs. arbitrary mappings), and (b) the effect of response modality, given high episodic demands (vocal/arbitrary vs. manual/arbitrary). Results clearly support a memory-based over a modality-specific explanation for the enhanced age effects in Stroop interference that were observed in the earlier experiments. Age and Stroop condition interacted with the memory contrast, while they failed to interact with the contrast that compared response modalities within arbitrary mappings. Under high memory load (i.e., with arbitrary mappings), there were relatively large age differences in the Stroop effect, regardless of the response modality. Again, this effect was particularly pronounced on trials on which the rule changed.

With Experiment 5 it was investigated whether (a) the result also holds with a minimal memory load, i.e., an arbitrary mapping consisting of only two stimulus-response rules, and whether (b) the result of small or absent Age×Stroop effects in the compatible condition is limited to vocal responses. A further topic of investigation was the degree to which verbal or spatial distractors differentially affect young and old adults. A spatial Stroop task was employed that allowed to compare compatible and arbitrary mappings using both vocal and manual responses. A minimal arbitrary mapping was implemented by pairing two target values with two response alternatives. Distractors were either arrows or words, which presumably use a spatial or phonological internal code, respectively. Both types of distractor were tested in the vocal and in the manual response modality, thereby providing conditions with high and low overlap of internal distractor and response codes. After it was found that there were problems with the particular choice of target-distractor ensembles in Experiment 5(a), the experiment was repeated with different ensembles as Experiment 5(b). Unfortunately, the new stimuli introduced their own problems. Because the conceptual design of the experiments was identical, data were pooled to increase power for the detection of consistent effects and at the same time discredit inconsistent (and consequently irrelevant) stimulus-related side effects. Due to the use of only two response alternatives, the magnitude of all effects was relatively small.[74] Even the minimal memory demands evoked by the two-element mapping rules caused an age difference in the Stroop effect to appear, whereas the size of the Stroop effect with compatible mappings was identical between age groups—a pattern that was also observed within each response modality. In contrast, response modality did not modulate age differences in the Stroop effect (although it did affect the size of the Stroop effect independent of age group, with a larger effect observed with vocal than with manual responses). Finally, the degree of overlap of distractor and response codes strongly influenced the size of the Stroop effect. Word distractors caused large Stroop effects with vocal responding, and weak Stroop effects with manual responding, while the converse pattern was obtained for arrow distractors. However, this effect of code compatibility was also independent of age. What did change with age was the relative influence that word and arrow distractors had on the size of the Stroop effect, independent of response modality. Words were much more effective distractors for old than for young adults, whereas arrow distractors caused similar Stroop effects in both age groups. Unfortunately, possible interactions of this effect with Mapping and Response modality could not be reasonably tested due to inconsistencies between Experiments 5(a) and (b).

Experiment 6 addressed the question whether there are age-differential strategies in the choice of internal codes used to represent the arbitrary mapping rules. For example, it is conceivable that old adults tend to rely on verbal codes, whereas young adults are able to flexibly choose the type of internal representation that is most appropriate in a given task context. A verbal coding strategy would be particularly susceptible to interference from word distractors, which were used exclusively in Experiments 2-4 and which, unlike arrow distractors, caused large age differences in the Stroop effect in Experiment 5. The consistently large Age×Stroop effects in the arbitrary mapping conditions might have at least partly been a consequence of the choice of internal code rather than of memory demands per se. In Experiment 6, different groups of participants were instructed to use either a spatial or a verbal representation of the mapping rules. Each group was tested with both a Stroop (color target, word distractor) and a reverse Stroop (word target, color distractor) task, with task order counterbalanced. Memory demands were kept at a relatively high level by using four manual response alternatives. An important result was that a large age difference in the Stroop effect was not only observed in the standard Stroop task, but also in the reverse Stroop task. Because the latter uses color distractors, it can be concluded that an age difference in the Stroop effect is not limited to verbal distractors. The instruction manipulation completely failed to influence responding on an overall level. This is difficult to interpret. It could mean that participants failed to follow instructions, that the kind of internal code does not matter in the tasks under investigation, that the manipulation was not strong enough, that verbal and spatial codes are used redundantly, etc. Yet a more detailed analysis revealed interesting effects of task order that might have hidden the instruction effects in the overall analysis. Even if task order was considered, effects of the instruction manipulation were limited to old adults. The instruction effect is most easily described if only the task performed first is considered. In the Stroop task, the spatially instructed group of old adults responded faster and produced a much smaller Stroop effect than the verbally instructed group. In the reverse Stroop task, the verbally instructed group of old adults was faster and produced a much smaller (reverse) Stroop effect than the spatially instructed group. If it is assumed that the effect is not due to individual differences, then the pattern might indicate that old adults profited from a guidance to use a task-appropriate code, whereas young adults were able to find the most appropriate on their own. Interestingly, the pattern was completely reversed in the tasks performed second. Although an explanation based on proactive task-set interference is available, results could

---

[74] This was expected based on the Hick-Hyman law.

also indicate that all of the instruction effects observed in the order analysis were merely caused by individual differences between groups of old adults. Thus the evidence that the choice of internal code used to represent mapping rules does affect old adults' performance is rather weak.

In summary, two outcomes were fairly consistent. First, there were small age effects with compatible, and more pronounced age effects with arbitrary S-R mappings. Second, age differences in the Stroop effect were relatively large with arbitrary target-response mappings, and absent with compatible mappings. Taken together, results show that age-related interference effects are amplified by episodic demands, supporting the interpretation that the reliability of episodic accumulators is reduced in old age. At an individual experiment level, results indicate that, in partial contrast to Verhaeghen and De Meersman's (1998) results, there are conditions under which an age-related inhibitory deficit can be revealed in Stroop-like tasks, namely if episodic task demands are sufficiently high. On the other hand, the results obtained in the compatible mapping conditions support Verhaeghen and De Meersman's conclusion that there is no age-related inhibitory deficit in Stroop tasks per se.

Graphical meta-analysis using Brinley Plots

Introduction

In the results reported thus far, in line with the majority of experimental cognitive aging researchers, an experimental approach was adopted. Basically, the strategy was to manipulate an independent variable that is assumed to reflect a cognitive construct, and to determine whether the manipulated variable interacts with age in the dependent measure.

If the age by condition interaction is significant, it is often concluded that there exists an age-related deficit in the cognitive mechanism underlying task performance. For example, if in a response-compatibility task one obtained higher age effects with arbitrary than with compatible stimulus-response mappings, one would conclude that memory is impaired in old age, assuming that memory demands lead to the cost introduced by arbitrary S-R mappings. Apart from difficulties related to ceiling and floor effects (Chapman & Chapman, 1973), the major problem with this interpretation is that is fails to take general slowing into account (for a more extensive critique of this research strategy see Salthouse, 1991). Because slowing is assumed to be proportional, general slowing predicts that any two conditions that differ in task difficulty (manifesting itself as an RT difference for young adults) lead to an age-by-condition interaction, given a sufficient sample size. Therefore, general slowing alone predicts over-additive age by condition interactions in an ANOVA of untransformed reaction time means. Is there anything to aid the researcher in determining whether the age-by-condition interaction she measured is overproportional, i.e. the degree of slowing is greater for one task condition than for the other? In the current report, several measures were taken to account for proportional slowing. First, all analyses were repeated with log-transformed reaction times, thereby switching into proportional measurement space. Second, the Stroop effect was normalized by an individual baseline latency, such as the reaction time in the neutral Stroop condition. In Experiment 6, it was attempted to statistically control for general slowing by taking into account a measure known to produce the general slowing pattern (digit-symbol scores). None of these control approaches seems completely satisfactory. For example, it has been shown that the analysis of log reaction times produces unbiased results only if the Brinley regression line goes through the origin (Faust et al., 1999), which is an unrealistic assumption. The normalization through division by baseline latencies appears to be justified only if these are uncontaminated by the effects under investigation. The statistical control approach assumes linearity, which might lead to a higher chance of finding significant residual interactions in tasks generating higher reaction times.

A different approach, and one that is more closely tied to the individual-differences research strategy, is to perform a Brinley plot meta-analysis. One problem here is that to obtain reliable regression estimates, one needs to measure a large number of conditions and a wide range of reaction times. Accordingly, Brinley analysis is most often not applicable to the interpretation of single experiments, but can only be applied if the domain under investigation has produced enough results.

Distinguishing between general and specific influences on cognitive age differences is an important issue in cognitive aging research. Brinley plots have become a standard tool for evaluating age differences, and the results from most graphical Brinley plot meta-analyses have been taken to support a general slowing model. Recall that in a Brinley plot, mean reaction time for old adults obtained in a number of studies and conditions are plotted as a function of the corresponding mean reaction times for young adults. Young adults' mean reaction times can be considered a measure of task difficulty or complexity. The analysis then proceeds by fitting a curve that best describes the relation between young and old adults' means. Typically, a linear regression is used for curve fitting, but other models such as a power function have also been used. The finding that the old-young relation is better described by two separate curves, one for process A and the second for process B, than by a single curve appears to be a clear signature of process-specific slowing.[75] However, separate curves are not often observed: a prototypical finding is the emergence of a single line with a slope greater than one and a negative intercept.

The fact that a difference in slopes has rarely been found has been taken as strong evidence for the generality of the slowing mechanism, although it has been

---

[75] Because only observable condition means are plotted, this reasoning implicitly assumes that tasks can be identified that selectively manipulate processes.

shown theoretically as well as in simulation studies that this conclusions is unjustified—a single slope does not imply that slowing is general in the sense that only a single mechanism is involved (e.g. Cerella, 1994; Dunn & Kirsner, 1988; Perfect, 1994). Conversely, if different regression slopes for different sets of experimental condition are found, then it is often theoretically interpreted to indicate a specific age deficit in processes underlying performance in the higher-slope conditions (e.g., Mayr et al., 1996; Verhaeghen et al., 2002; Hale & Myerson, 1996; Fisk et al., 1990). Again, objections to this conclusions have been voiced. However, whereas separate slopes might not strictly imply different slowing factors (Salthouse, 1987) separation of slopes at least are only reconcilable with general slowing if rather strong processing assumptions are made. Although the finding of a difference in slopes might not completely rule out general slowing, at least it does make it less likely an explanation. In fact, if nonmonoticity is found in a Brinley plot, then it can be concluded under relatively weak additional assumptions that separate mechanisms are involved (Cerella, 1994; Dunn & Kirsner, 1988). In summary, the finding of a single slope is less diagnostic than a finding of separate slopes. Although Brinley plots have less power than ANOVA to detect age differences, the analysis in itself can produce interesting results. Its main advantage is that it can provide additional evidence for the fact that not all of the between-condition variance can be explained by general slowing. If different Brinley slopes are found for different conditions, a general slowing explanation is hard to defend.

In the current work, Experiments 2-6 (described above) investigated age differences in the Stroop effect. Because these data were all collected in the same lab, i.e., there is relatively little between-experiment variance in environmental conditions, we have a data set that is well suited for a graphical meta analysis using Brinley plots. In most experiments, aspects of stimulus-response compatibility were varied in addition to Stroop condition. How do the experimental results transfer to the Brinley plot? Does the consistency of the Age×Mapping×Stroop condition interaction mean that regression lines with separate slopes connect points from compatible and arbitrary mapping conditions in Brinley space? The argument made in the following paragraphs shows that this is not the case. Rather, different slopes are only predicted if the intercepts are allowed to vary between experiments.

Only experiments that manipulated Stroop condition were included in the Brinley analysis[76]. Except for Experiment 6, all of the included experiment additionally manipulated stimulus-response compatibility. Experiment 6 has a special status also because the proportion of incongruent trials was lower than in the other experiments. For all experiments, the condition means of old adults were regressed on the condition means of young adults, and it was tested whether slopes, intercepts, or both were influenced by Stroop condition and mapping. The test for parameter differences followed the procedure suggested by Verhaeghen et al. (2002).

Figure 18 shows two thirds of the data points included in the meta-analysis, namely the condition means for young and old adults in the congruent and incongruent Stroop condition (i.e., the neutral conditions are not plotted, but included in the meta-analysis). Congruent and incongruent points from otherwise corresponding conditions in single experiments are joined by a line, and line color indicates whether the condition involved compatible or arbitrary mappings. The lines thus represent the Stroop effect in Brinley space.

Before the Brinley analysis is presented, let us briefly consider the predictions made by (a) a prototypical general slowing model, and (b) by a model that assumes a specific age deficit in episodic accumulators.

General Slowing Predictions

The finding that a single linear function can describe the relation between young and old adults' performance over a range of task complexity has been taken as evidence for general slowing. Cerella (1990) has formulated a processing model of general slowing that predicts linear Brinley plot. He conceived of information processing in a task as finding the shortest route through a network of connected neurons, and of age effects as breaking links in the networks, thus requiring diversions. General slowing leads to the deceptively simple model

$$\begin{aligned} RT_{young} &= S + C_1 + C_2 + \cdots + C_n \\ &= S + \sum_{i=0}^{n} C_i \quad . \\ RT_{old} &= S + \lambda C_1 + \lambda C_2 + \cdots + \lambda C_n \\ &= S + \lambda \sum_{i=0}^{n} C_i \quad , \end{aligned}$$

where S designates processing time for the sensorimotor stage, and C designates processing times in consecutive serial cognitive stages. The Brinley function

$$RT_{old} = \beta_0 + \beta_1 RT_{young}$$

can be obtained from this model by expressing old RT as a linear function of young RT, i.e.

$$RT_{old} = (1 - \lambda)S + \lambda RT_{young} \quad .$$

Because stages are reached in succession, and all cognitive stages have the same slowing factor, separate slopes cannot be obtained for separate tasks.

Episodic Slowing Predictions

The episodic slowing model makes fairly coarse predictions insofar as only three cognitive stages, C_1, C_2, and C_3 are distinguished, in addition to a sensorimotor stage for which slowing is assumed to be absent (based on Cerella's (1985) data, more realistic would be a slowing factor of about 1.2 for the sensorimotor stage). It is

---

[76] This amounts to an exclusion of Experiment 1.

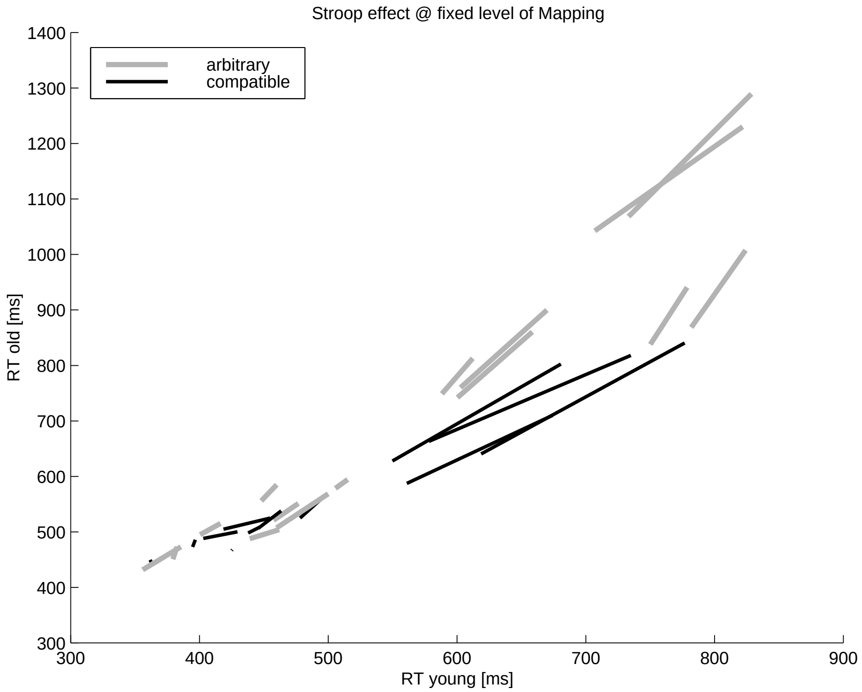


Figure 18. Young and old adults' mean reaction times in incongruent and congruent response change conditions of Experiments 2-6. Lines connect incongruent and congruent Stroop conditions at corresponding mapping conditions per experiment.

not a processing model (although it could be extended to become one, for an attempt see the model sketched on pp. 27 ff.), but it is structural in the sense that it describes the relation between the outcomes of processes. Let us call the three cognitive stages 'stimulus classification' (C_1), 'episodic processing' (C_2), and 'response preparation' (C_3). The core assumptions are that difficulty effects arising in stage C_1 are amplified by stage C_2, and that there exist conditions under which stage C_2 can be bypassed. The amplification assumption can be formulated by regarding C_2 duration a function of C_1 duration. The functional form of this relation has to be specified—for simplicity, I chose a multiplicative function:

$$C_2 = f(C_1) = b_1 C_1 \quad .$$

The slowing factors for C_1, C_2, and C_3 can be different. The model thus consists of a set of equations:

First, for conditions where C_2 can be bypassed

$$\begin{aligned} RT_{young} &= S + C_1 + C_3 \quad . \qquad (14)\\ RT_{old} &= S + \lambda_1 C_1 + \lambda_3 C_3 \quad . \end{aligned}$$

Second, for conditions where C_2 cannot be bypassed

$$\begin{aligned} RT_{young} &= S + C_1 + C_2 + C_3 \qquad (15)\\ &= S + C_1 + b_1 C_1 + C_3\\ &= S + (1 + b_1) C_1 + C_3 \quad .\\ RT_{old} &= S + \lambda_1 C_1 + \lambda_2 C_2 + \lambda_3 C_3\\ &= S + \lambda_1 C_1 + \lambda_2 b_1 C_1 + \lambda_3 C_3\\ &= S + (\lambda_1 + \lambda_2 b_1) C_1 + \lambda_3 C_3 \quad . \end{aligned}$$

Here, the λ_i are slowing factors of the associated stages C_i, with $i \in \{1 \ldots 3\}$, and b_1 describes the dependency of C_2 duration on C_1 duration.

For a manipulation of C_1, at a given level of C_3, the predicted Brinley slope for the first set of equations, (14), will be influenced only by λ_1, and the Brinley function given by

$$RT_{old} = (1 - \lambda_1) S + \lambda_1 RT_{young} \quad ,$$

with a slope of

$$\beta_1 = \lambda_1 \quad , \qquad (16)$$

and an intercept of

$$\beta_0 = (1 - \lambda_1) \quad . \qquad (17)$$

The same also holds between compatible experimental conditions, if it is assumed that the slowing factors of early cognitive processing and of response preparation are the same, $\lambda_1 = \lambda_3$.

However, the Brinley slope for the second set of equations, (15), will be a function of λ_1, λ_2, and b_1. For a given task (response condition), the Brinley function for the arbitrary mapping condition has a slope of

$$\beta_1 = \frac{\lambda_1 + \lambda_2 b_1}{1 + b_1} \quad , \qquad (18)$$

and an intercept of

$$\beta_0 = S + \lambda_1 C_1 + \lambda_3 C_3 - (S + C_1 + C_3) \frac{\lambda_1 + \lambda_2 b_1}{1 + b_1} \quad . \qquad (19)$$

Between tasks or response conditions, the situation is even worse. The slope of the global Brinley function will depend not only on the 'episodic' and 'early cognitive' slowing factors, but also on the sizes of the between-experiment effects elicited by changes in response conditions (see Appendix, equation (29)), meaning that the parameters of a single regression line carry no useful interpretation. An attempt at a mathematical derivation of the predictions is given in Appendix I (pp. 109 ff.). There, the predictions are also extended to compare conditions differing in demands on response preparation, e.g., vocal and manual responding.

To the degree that λ_1 and λ_2 are different, different slopes are predicted for the two sets of equations. Because we assume that stage C_2 can be bypassed in the compatible mapping conditions ($b_1 = 0$), we predict that the slope of the lines connecting groups means for congruent and incongruent Stroop conditions in tasks (a) without and (b) with arbitrary mappings will be different (assuming that $\lambda_1 \neq \lambda_2$).

One problem is that while within a given task and for a given level of C_2, one can assume that C_3 and b_1 are constant, this might not be the case between tasks. In particular, it is likely that task complexity as measured by young adults' reaction times is, among other factors, a function of memory demands. Thus between tasks differing e.g. in memory demands, b_1 may change. However, for simplicity, we assume that b_1 is constant across tasks, and only the 'setup cost', C_3, may change between experiments. For example, finding a response alternative among four keys is easier than finding a response alternative in the potentially very large set of male first names. We assume that in the latter case, C_3 might be larger than in the former, thus the C_3 in a manual response task might be different from C_3 in a task with male first name responses. However, this does not further amplify C_1 effects.

The episodic buffer model makes specific predictions with respect to the Brinley plot. First, a single line is predicted for compatible mapping conditions. Second, slopes for arbitrary mapping conditions in each comparable task should be the same. Third, the intercept for the regression lines can vary between experiments—thus for arbitrary mappings, a number of parallel lines is predicted, one for each experiment. To understand why the model makes these predictions, we have to reconsider the model described above.

For compatible mappings only one stage is needed because by assumption the episodic accumulator stage can be bypassed. The regression slope is determined by a single slowing factor, which reflects slowing in the process tapped by the task. Since across tasks the same 'early difficulty' manipulation was used, namely Stroop condition, the processes tapped by the task can be assumed to be approximately constant across experiments. The intercept of the compatible-condition regression line is a function of the slowing factor and the duration of peripheral processes. The latter is assumed to be constant across experiments—hence for compatible mapping conditions, neither the intercept nor the slope of the Brinley regression should depend on the experiment, i.e. points from all experiments should fall on a single line, which is the first prediction.

The within-experiment Brinley slope for the arbitrary mapping conditions depends on both the slowing of the first stage and the slowing of the second, 'episodic' stage, as well as on the degree of the dependency of the second on the first stage. The first stage slowing factor is simply the slowing factor of the processes tapped by the Stroop task, i.e. the slope of the regression for compatible conditions. The second stage slowing factor is slowing of the 'episodic' stage, implementing arbitrary mappings. Because this stage is supposedly the same across experiments, the second slowing factor should not change between experiments. First and second stage slowing factors are averaged, with the second stage slowing factor weighted by an 'amplification factor' b_1 describing the degree of dependency of second on first stage duration, to obtain the within-experiment arbitrary-condition Brinley slope. It is not clear whether this factor is constant across experiments, but let us assume that it is approximately so, because the Stroop manipulations were comparable–thus we arrive at the second prediction from the model, approximately equal slopes for the arbitrary mapping conditions across experiments.

The third prediction is the most unusual in the context of Brinley plots. Usually, only one intercept parameter per meta-analytic factor is allowed to vary freely. However, the interactive experimental designs used, combined with the two-stage slowing model, strongly suggest that a separate intercept should be fitted per experiment. This is because the intercept parameter is a complex function of the duration of the peripheral and both central stages, the corresponding slowing factors, and the 'amplification factor'. In particular, it is also a function of the 'response preparation' stage duration, C_3, which is assumed to be constant for a given arbitrary mapping condition, but can vary between experiments (and response conditions).

Data selection

Because of the strong response repetition effects observed with vocal-arbitrary mappings, only data from response change conditions were fitted to reduce error variance. In an initial attempt at fitting the data, two data points were identified as clear outliers in a preliminary analysis, by both Cook's Distance and Mahalanobis distance. These outliers were located far above the diagonal, thus indicating an exceptionally large age difference. Because both of these came from the incongruent color naming condition of Experiment 6, which was not well comparable to the other incongruent conditions due to the low proportion of incongruent trials (see above), it was decided to leave these out of all analyses (alternatively, a separate manual-arbitrary slope parameter could have been fitted for Experiment 6). Lastly, compatible-manual conditions were only realized in Experiment 5, where overall reaction time was low. Compatible-vocal conditions originated from several experiments, covering a wider range of young RT. I decided to keep compatible-manual data in the analysis, but not to treat them separately from compatible-vocal.[77] The reader should keep in mind that generalizing from the present data to manual-compatible conditions with a higher degree of difficulty might be problematic. After data selection, the data available for Brinley analysis consisted of a total of 85 points describing pairs of young-old condition means obtained in 5 experiments.

Data Analysis

Because the experiments were designed to test the episodic slowing model against a general slowing model, several alternative regression models were formulated and tested. We started with a simple 'general slowing' regression model, assuming that a single line is sufficient to fit the data. The model was then extended (a) to include a separate line for arbitrary mapping conditions, and (b) to include separate intercepts for arbitrary mapping conditions in each experiment. The extended model (a) is a standard approach in Brinley analysis to test whether different conditions differ in their influence on aging (e.g., Verhaeghen et al., 2002). Here it was tested whether type of mapping had a significant influence on the parameters of the regression. To my knowledge, this approach is merely descriptive and not guided by predictions from a theoretical slowing model. Extended model (b) was guided by the structure of the 'episodic slowing' model, which makes specific predictions with respect to the Brinley plot. As discussed above, it predicts a single line for the compatible conditions, and a set of lines with equal slopes, but possibly separate intercepts for the arbitrary conditions. Hence in the regression analysis, the intercepts for the arbitrary mapping Brinley functions were allowed to vary. Additionally, separate intercept parameters were allowed within Experiments 4 and 5 to distinguish between manual and vocal arbitrary mappings, because there might be possible differences in setup costs. Taken together, the multiple-intercept model (b) predicts old condition means by incorporating some, but not all available information about young condition means, type of mapping, and experiment, The linear model has two free parameters for the compatible conditions, one for the intercept and one for the slope, one slope parameter for the arbitrary mapping conditions, and $1+1+2+2+1=7$ intercept parameters for arbitrary conditions in Experiments 2 through 6, respectively. Only the two slope parameters make direct use of young RT.

Tests were performed within the General Linear Model approach, which offers an elegant way of framing separate regression models within a superordinate model. To test model predictions, we used a General Linear Model with the dummy-coded categorical predictor Mapping, and the continuous predictor young RT.

In the General Linear Model, all regression models have the common form of $y = X\beta + e$, where y stands for old RT, X is the design matrix that specifies the predictors including young RT, β is a vector of regression weights to be estimated, and e is an error term. The estimates are obtained by solving for β, which is given by the well-known solution $\beta = (X'X)^{-1}X'y$ if the design matrix is of full rank. We did not fit nonlinear models mainly because we lack a theoretical motivation for doing so.[78]

The different models were expressed by using different design matrices X. For the simplest general slowing model, X had only two columns, one column of ones to estimate the intercept, and one column of young RT condition means to estimate the slope. For the two-line model, the design matrix included an indicator variable that coded for arbitrary mapping (i.e. it was set at 1 for arbitrary mapping conditions and at 0 otherwise), and a fourth column that was the product of the third column and young RT. In this model, the parameter β_1 for the first column estimates the intercept for compatible mapping conditions, and β_2 gives the slope of the line for compatible conditions. Parameters β_3 and β_4 estimate the difference in intercept and slope relative to β_1 and β_2, respectively, so that the intercept and slope of the regression line for arbitrary mappings is given by $\beta_1 + \beta_3$ and $\beta_2 + \beta_4$, respectively. Finally, the multiple-intercept model was like the two-line model with the third column dropped, and replaced by additional indicator variables. For each arbitrary mapping condition

---

[77] It could be argued that not allowing for an extra intercept for the manual-compatible condition leads to biased estimates of the regression parameters for the compatible conditions. This is because only Experiment 5 had a manual-compatible condition, and overall 'complexity' (reaction time) in Experiment 5 was rather low. If any complexity-related between-experiment variable interacted with the slope for the manual-compatible function, we would have missed it in our realizations of the 'random variable' experiments. Empirical results (e.g., the positively accelerated power function described by Hale et al., 1987) might lead to the speculation that across meta-analyses, Brinley slopes tend to be smaller at low complexity levels, so that inclusion of a condition at an overall low level of complexity might bias the regression slope towards one and the intercept towards zero. To evaluate whether manual-compatible data points have to be treated separately, two alternative approaches were also used, namely (a) to drop manual-compatible points, and (b) to estimate separate intercept and slope parameters for manual-compatible points. Estimates for the (vocal-)compatible condition in (a) did not differ much from estimates in the models described in more detail below. Parameter estimates for the manual-compatible condition in (b) did not differ significantly from vocal-compatible parameters. Thus it appears to be justified to leave the manual-compatible conditions in the analysis, and not to treat them separately from the vocal-compatible conditions.

[78] There exists one prominent model of age-related slowing that predicts a power-law relationship of young and old RT in the Brinley plot (Myerson et al., 1990). In an initial explorative analysis using nonlinear regression of old RT on the predictors of age, mapping, and response modality, we found that exponential functions fit our data best, while the fits of linear and power functions were not distinguishable from each other. However, the multiple-intercept linear regression model described below fits the data better than the exponential model, and it is theoretically motivated. I therefore do not present details of the explorative exponential fit here.

in each experiment, a new indicator variable was added that was set at one for the corresponding condition (e.g. Experiment 2, manual-arbitrary), and at zero otherwise. The dummy variable coding for the change in slope in the arbitrary conditions (i.e. column 4 in dual-intercepts model, column 10 in the design matrix below) was left unchanged. Here is an example of different rows of the design matrix in the multiple-intercepts model:

$$X = \begin{bmatrix} 1 & y & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & y & 1 & 0 & 0 & 0 & 0 & 0 & 0 & y \\ 1 & y & 0 & 1 & 0 & 0 & 0 & 0 & 0 & y \\ 1 & y & 0 & 0 & 1 & 0 & 0 & 0 & 0 & y \\ 1 & y & 0 & 0 & 0 & 1 & 0 & 0 & 0 & y \\ 1 & y & 0 & 0 & 0 & 0 & 1 & 0 & 0 & y \\ 1 & y & 0 & 0 & 0 & 0 & 0 & 1 & 0 & y \\ 1 & y & 0 & 0 & 0 & 0 & 0 & 0 & 1 & y \end{bmatrix}$$

Results

In a first step, a general slowing model was fitted, that estimated old RT purely on the basis of young RT. Compared to published “general slowing” Brinley plots, which often have R squared in the region of .95, the fit of this null model was mediocre, $R^2 = .865$, $RMS_e = 82.33$. The ANOVA table for general slowing model gives $F(2,83) = 534.04$, $p < .001$, $MS_e = 6778.336$. Thus as expected, in comparisons to a null model that predicts a constant (mean) old RT, young RT is a valuable predictor—old RT covaries with young RT. The estimated regression equation is

$$RT_{old} = -124.90 + 1.45 RT_{young} \quad .$$

Both the slope (1.451, $s.e. = .063$, $t = 23.11$, $p < .001$) and the intercept (-124.899, $s.e. = 35.49$, $t = -3.52$, $p = .001$) were in the range typically observed and significantly different from zero.

In a second step, a standard “Brinley analysis” was performed that simply added an indicator variable to estimate the intercept for arbitrary mappings, and a product of the indicator variable and young RT to estimate the associated slope. This model estimates two regression lines, one for compatible mappings, $(RT_{old})_c = \beta_{1c} + \beta_{2c}(RT_{young})_c$, and one for arbitrary mappings, $(RT_{old})_a = \beta_{1c} + \beta_{1a} + (\beta_{1c} + \beta_{1a})(RT_{young})_a$. Inclusion of the contrasts for arbitrary mapping conditions significantly improved model fit, $F_{Change}(2,81) = 24.04$, $p < .001$. The new model had an R^2 of .916, but the standard error of the estimate was still rather high, $RMS_e = 66.02$. For compatible mappings, the estimated slope parameter, .958 ($s.e. = 095$), did not significantly differ from 1, and the estimated intercept, 87.534 ($s.e. = 49.167$), was not significantly different from zero, $t = 1.78, p = .079$. The estimated *increase* in slope for the arbitrary mappings was .616 ($s.e. = .114$), which differed from zero, $t = 5.40, p < .001$. Hence the slopes for compatible and arbitrary mappings (.958 vs. 1.574, respectively) were significantly different. The estimated *decrease* in intercept associated with arbitrary mappings was 258.074 ($s.e. = 61.835$) and differed significantly from zero ($t = -4.17, p < .001$). The estimated parameters lead to the following equations describing the best-fitting straight lines, for compatible mappings:

$$(RT_{old})_c = 87.53 + .96(RT_{young})_c \quad ,$$

and for the arbitrary mappings:

$$(RT_{old})_a = -170.54 + 1.57(RT_{young})_a \quad .$$

Thus if results are interpreted in the standard way, slowing in tasks with arbitrary mappings is a fact, while there is no slowing in tasks with compatible mappings. However, this interpretation appears flawed, because between-experiment effects that were not directly related to the constructs under investigation are confounded with the effects of interest, as has been pointed out by Sliwinski and Hall (1998) and will be elaborated on below. This can be easily demonstrated if Experiment 6 is dropped from the analysis. If the separate-lines model is fitted to the data from Experiment 2-5, the parameters for the arbitrary mapping change dramatically (because there was no compatible mapping condition in Experiment 6, parameters for the compatible mapping did not change). The slopes for arbitrary and compatible mappings in the new model are still significantly different, however the new estimated regression for arbitrary mappings is $(RT_{old})_a = -3.06 + 1.21(RT_{young})_a$, which is drastically different from the function reported just above. Does this mean that slowing due to arbitrary mappings is almost negligible in Experiments 2 to 5, and pronounced in Experiment 6? I think that the explanation is different, and is related to Sliwinski and Hall’s argument.

To control for the effects of experiment that were not directly related to the variables under investigation, I allowed for separate intercepts in the arbitrary mapping conditions, while still constraining the regression model to a single slope parameter for the arbitrary mappings. This gives us the regression model predicted by the theoretical formulation of the “episodic two-stage slowing model”, which allows for a part of the second stage to vary independently of the duration of the first stage. The duration of this independent part is allowed to change between experiments, because it seems plausible that the differences between experiments somehow affect later cognitive stages. Thus, to test predictions from the theoretical model, as well as to control for theoretically uninteresting between-experiment effects, in a next step, a set of indicator variables for intercepts in the arbitrary mapping conditions were added, consisting of separate variables for each combination of experiment and response modality. Figure 19 shows the results of the dual-slope, multiple-intercept regression analysis, overlaid on the Brinley plot data. Note how the pattern of fitted regression lines resembles the empirical pattern shown in Figure 18 (p. 89).

Adding the indicator variables and the mapping effect to the equation significantly improved model fit, in comparison to the general slowing model, $F_{Change}(8,75) = 124.24$, $p < .001$, as well as in comparison to the standard two-line Brinley model, $F_{Change}(6,75) = 99.30$,

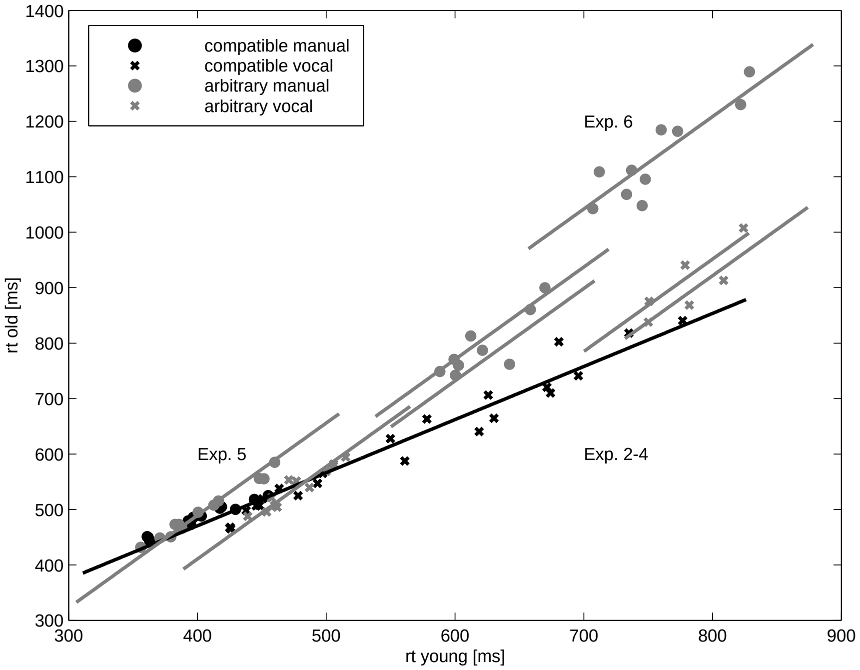


Figure 19. Brinley plot data, with regression lines fitted by the dual-slope, multiple-intercept model overlaid.

$p < .001$, so that the extended model gave a reasonably good fit to the data, $R^2 = .991$, $RMS_e = 22.94$; $F(9,75) = 874.63$, $p < .001$, $MS_e = 526.33$. Admittedly, this comes at a cost of 6 additional parameters in comparison to the two-line model, however, none of these parameters is of central theoretical interest. What does matter is the estimate of the arbitrary slope parameter, which is indicative of age-related amplification of early (Stroop) difficulty effects by episodic task demands. Table 9 lists the estimated parameters for the extended model. For easier interpretation, the last column presents estimates that are rescaled to reflect absolute terms instead of deviations from a reference category.

We note two interesting points. First. the slope of the regression for the compatible mapping condition (0.96) is not significantly different from one, and the intercept (87.53) is positive[79]. Thus although the individual data points on this line stemmed from congruent and incongruent conditions of Stroop experiments, slowing was absent. Apparently, with compatible mappings, the processes that lead to Stroop interference are not affected by aging. Second, the slope for the arbitrary mapping conditions (1.67) was significantly different from the compatible slope ($t = 6.07, p < .001$), and significantly different from one ($t = 5.71, p < .001$), although again points along the slope originated from congruent and incongruent Stroop conditions. Taken together, this is indicative of amplification of age differences in the Stroop effect when arbitrary stimulus-response mappings are used. This result supports the interpretation that even in apparently trivial reaction time tasks, age differences are primarily caused by memory demands.[80] A third point to note is that the intercept parameters for arbitrary-vocal mappings are larger than for arbitrary-manual mappings. This points to the fact that arbitrary-vocal mappings impose an extra cost for both young and old adults. This might have been caused by the fact that in contrast to manual responding, no external representation of the response set was available in the arbitrary-vocal conditions. The fact that an extra setup cost might lead to a smaller degree of cascaded processing was already discussed in the individual experiment sections.

To see whether the arbitrary mapping slope estimate from the multiple-intercepts analysis is a better indicator of the true slowing factor in arbitrary mapping conditions than the slope estimated by the standard two-line regression model, the analysis was again repeated without the data points from Experiment 6. The slope estimate for the arbitrary mapping condition in Experiments 2-6 was now 1.596. Thus now the estimates were comparable: whether or not Experiment 6 was included, the old-young slope estimate for Stroop effects under arbitrary mapping conditions is about 1.6. Compare this result with that from the standard Brinley analysis, where eliminating Experiment 6 data drastically

---

[79] If an extra intercept parameter for compatible-manual conditions is added, leading to a regression RMS_e of 21.95, then the compatible-vocal intercept (44.21) no longer differs from zero—however, the compatible-manual intercept (71.65) does.

[80] In fact, if we assume that our model (15) is valid, we can estimate that the slowing factor caused by the arbitrary mapping is at least 1.67. The results of our Brinley analysis give us $\beta_1 = 1.67$ for the arbitrary mappings, and an estimate of $\lambda_1 \approx 1.0$ that was obtained from the compatible mappings. Although we do not know either b_1 or λ_2 in equation (18), we do know that $\lambda_2 > \beta_1$ (since $b_1 > 0$), and can rearrange terms to obtain $\lambda_2 = \beta_1 + (\beta_1 - \lambda_1)b_1$, here, $\lambda_2 = 1.67 + .67/b_1$. If the degree of dependency of C_2 duration on C_1 duration is very high, then $\lambda_2 \rightarrow 1.67$. If the dependency is smaller, then λ_2 might be much larger.

Table 9
Parameters for the multiple-intercept, dual-slope model. The column labeled "B" lists the parameter estimates, relative to a reference category. The last column presents the linear combinations of parameters used to determine intercept and slope for the regression lines. For example, $RT_{old} = -267.81 + 1.67\,RT_{young}$ is the equation describing the line for arbitrary-manual mappings in experiment 4. Rows 4-10 contain statistics for the arbitrary mapping intercepts, which were allowed to vary between experiments and response conditions.

					95% CI for B		absolute
Parameter	B	s.e.	t	sig. (p)	lower	upper	Int./Slope
Slopes							
compatible	0.96	0.03	28.870	0.000	0.89	1.02	0.96
arbitrary	0.71	0.12	6.070	0.000	0.48	0.94	1.67
Intercepts							
compatible	87.53	17.09	5.123	0.000	53.50	121.57	87.53
arbitrary manual, Exp. 2	-316.65	71.61	-4.422	0.000	-459.29	-174.00	-229.12
arbitrary manual, Exp. 4	-355.34	74.15	-4.792	0.000	-503.06	-207.62	-267.81
arbitrary manual, Exp. 5	-264.66	48.82	-5.421	0.000	-361.92	-167.41	-177.13
arbitrary manual, Exp. 6	-211.94	86.69	-2.445	0.017	-384.65	-39.24	-124.41
arbitrary vocal, Exp. 3	-498.90	92.65	-5.385	0.000	-683.46	-314.33	-411.36
arbitrary vocal, Exp. 4	-468.92	87.75	-5.344	0.000	-643.73	-294.11	-381.38
arbitrary vocal, Exp. 5	-343.31	56.53	-6.073	0.000	-455.92	-230.70	-255.78

changed the estimate for the arbitrary mapping slope. Controlling for between-experiment variance by allowing per-experiment intercept parameters led to much cleaner and more robust estimates of slowing due to arbitrary mappings.

To check whether a single slope parameter was sufficient to capture the essence of the slowing functions, subsequent analyses were performed that loosened the restrictions on the slope parameter. First, the arbitrary slope parameter was dropped, and replaced by separate slope parameters for manual-arbitrary and vocal-arbitrary mappings. Second, separate slopes were allowed for each condition for which an intercept was estimated, i.e., for each combination of experiment, arbitrary mapping, and response modality. Allowing separate slopes for the manual- and vocal-arbitrary conditions did not change R squared ($R^2 = .991$), and led to an increase in RMS_e (23.08). Furthermore, the estimated slopes for the arbitrary-manual and arbitrary-vocal conditions were very similar, and the 95%-confidence intervals overlapped.[81] Thus, we can conclude that introduction of separate slowing estimates for manual- and vocal-arbitrary conditions was unnecessary. There remains the possibility that the effect of response modalities was obscured by aggregation across experiments. To test this, in addition to separate intercepts, we also allowed for separate slope estimates for each response modality in the arbitrary condition of each experiment. The fit of the model ($R^2 = .992$, $RMS_e = 22.55$) was slightly better than with a single slope for arbitrary mappings, but at the cost of seven additional parameters. All the estimated slope parameters for arbitrary mappings were different from one, with a range between 1.41 and 3.12 (rescaled), and most were significantly so. However, the confidence intervals overlapped in all pairwise comparisons of parameters. Taking into account the large variability in parameters, this indicates that the estimates were not very reliable, which comes to no surprise, because often only three points contributed to an estimate. An alternative way of comparing this model with the original model presented in Table 9 is to compare the F_{Change} relative to their parent model, which consists of a constant, a slope for young RT, and intercept parameters for all arbitrary conditions. The relevant values were $F_{Change}(7,68) = 6.68$, $p < .001$, for the present, exhaustive model, and $F_{Change}(1,75) = 36.84$ for the model with a single arbitrary slope parameter. Thus, we conclude that the more parsimonious model with a single slope parameter for the arbitrary mapping ought to be preferred.

Another approach to control for between-experiment differences in factors that are not of primary theoretical interest is to perform a Brinley analysis of Stroop effects. Because the Stroop effect is a difference measure, the setup costs common to congruent and incongruent conditions cancel out. All experiments that were included in the previous analysis manipulated Stroop condition, so we can calculate the Stroop effect as the reaction time difference between incongruent and congruent condition means per age group. This was done for each combination of experiment, mapping, and response modality, and the resulting Stroop effects were submitted to Brinley analysis. Graphically, the results are presented in Figure 20. Clearly, inspection of the plot suggests that the estimated slopes for compatible and arbitrary mappings are different. The regression analysis confirms this impression, as can be seen in Table 10. The overall fit for the combined regression estimating two intercepts and two slopes at the same time is not too bad, considering the fact that notoriously unreliable difference values were analyzed, $R^2 = .899$, $RMS_e =$

---

[81] Here, the parameter estimates the increase in slope compared to compatible-vocal, values are in the format *estimate-1.96 s.e.* $\leq$ *estimate* $\leq$ *estimate+1.96 s.e.*: The change in slope is $.44 \leq .70 \leq .96$ for the *arbitrary-manual* conditions, and $.24 \leq .76 \leq 1.28$ for the *arbitrary-vocal* conditions.

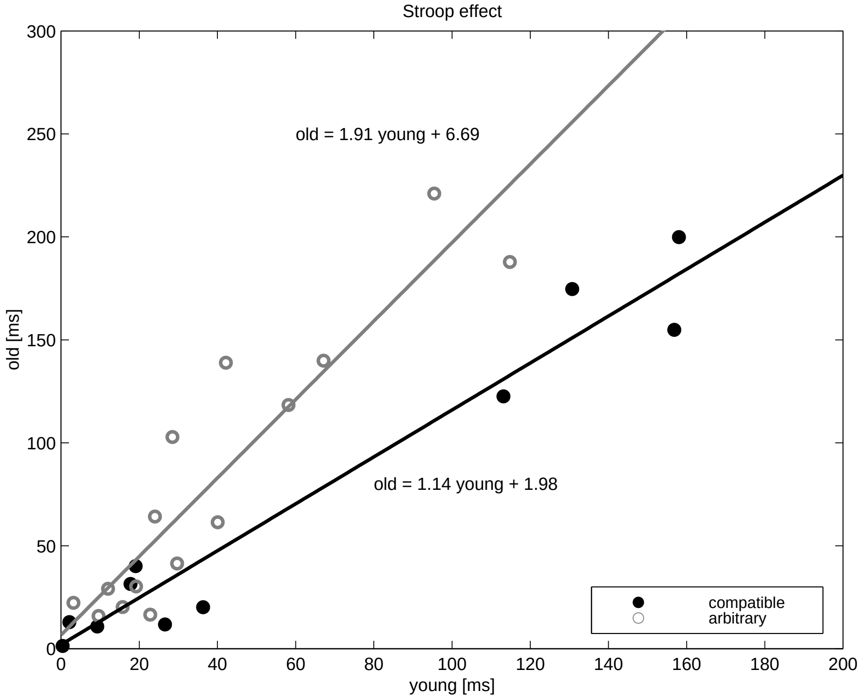


Figure 20. Brinley plot of the Stroop effect difference measure. Separate regression lines were fitted for compatible and arbitrary mappings.

23.59; $F(3,23) = 67.60$, $p < .001$, $MS_e = 556.46$. Neither of the intercepts is significantly different from zero, which was to be expected because peripheral processing times fell out of the analysis of RT differences. The slope for the arbitrary mapping conditions is significantly higher than for the compatible conditions, while the latter is not significantly different from one. The arbitrary slope in the analysis of mean reaction times reported above was somewhat lower than in the analysis of Stroop effects, which might be due to the fact that data from the neutral condition were included in the former, but not in the latter analysis.

Discussion

In summary, results are straightforward. Stroop manipulations do not appear to cause age differences in reaction time when the stimulus-response mapping is compatible. They do cause age differences, with a slowing factor of about $1.6 - 1.9$, when the mapping is arbitrary. The first finding is in accord with meta-analytic results obtained by Verhaeghen and De Meersman (1998), who found no specific age-related slowing in processes contributing to Stroop interference. These authors inspected several possible moderator variables, all of which were found to be irrelevant. In contrast, episodic task demands were found to be a strong moderator of Brinley slopes for Age × Stroop effects in the present analysis. This was not only true for the somewhat unorthodox (but theoretically justified) multiple-intercepts model, but also in a more traditional analysis restricted to single lines for each level of the moderator variable (the "standard two-line model").

A common results of both Verhaeghen and De Meersman (1998) and the present analysis is that Stroop condition itself is not a moderator of the age × complexity effect. In the present analysis, Brinley regression slopes at each level of episodic difficulty connected points from congruent and incongruent Stroop conditions. For at least some of the arbitrary mapping conditions, points originating from congruent Stroop conditions were located above the regression line for the compatible mappings. An approach that would have tried to fit regression lines based exclusively on knowledge about Stroop condition would not have led to a satisfactory fit. However, the slope connecting points from congruent, neutral, and incongruent Stroop conditions in the arbitrary mapping conditions was similar in all experiments, and different from the compatible condition slope. If Stroop-like interference acts at the stage of memory retrieval, then age effects become relatively large.

Interestingly, the Brinley slope for the compatible mapping conditions was much shallower than the 1.88 slope reported by Verhaeghen and De Meersman. This is surprising, because all of the studies included in their meta-analysis used a vocal response format with a compatible concept-response mapping. While it is not clear what accounts for this difference in slope between the published and the current meta-analyses, one reasons appears plausible. The critical factor might be fact that the number of trials in the present experiments was several times as large as in the studies included in Verhaeghen and De Meersman (1998), where scores were based on less than 50 incongruent items on average. Most of the studies in that meta-analysis used a blocked presentation format (most often the Stroop

Table 10
Parameters for the regression of old adults' on young adults' Stroop effects, with mapping as a factor.

Parameter	B	s.e.	t	sig. (p)	95% CI for B lower	95% CI for B upper	absolute B rescaled
compatible slope	1.14	0.11	10.240	0.000	0.91	1.37	1.14
arbitrary slope	0.77	0.23	3.407	0.002	0.30	1.23	1.91
compatible intercept	1.98	9.21	0.215	0.832	-17.08	21.04	1.98
arbitrary intercept	4.71	13.40	0.352	0.728	-23.00	32.43	6.69

Color-Word test from neuropsychological test batteries), where about 30 stimulus ensembles are presented simultaneously on a sheet of paper.[82] It has been shown that performance in the Stroop task improves with practice (Dulaney & Rogers, 1994), and that the absolute improvement is larger for old than for young adults, while the proportional improvement (compared to a pre-test color-naming baseline latency) is identical in both age groups.[83] Dulaney and Rogers have argued that most of the improvement can be attributed to general, task-related factors such as general response and scanning strategies, rather than to the development of a specific automatic reading-suppression response. For example, if a blocked presentation format is used, the scanning strategy is subject to improvement. On the other hand, if a task with the blocked presentation format uses only a single or a few cards, then the scanning strategy might contribute more to the age differences than the actual interference process. A Brinley plot of mean latencies from Stroop experiments involving a large number of trials will likely show a shallower slope, because of a greater amount of practice with general task factors. The Brinley slope in the Verhaeghen and De Meersman (1998) meta-analysis of Stroop tasks with compatible mappings was rather steep and in fact closely resembles the slope I obtained in the arbitrary mapping conditions. This might indicate that a large amount of the age-related variance between the studies included in the meta-analysis was due to executive, strategic, or general task-related factors.

Methodological aspects

A point that needs discussion is the approach to estimate separate intercepts in the arbitrary mapping conditions for each experiment and response modality. In my opinion, separate intercepts per experiment indicate complexity effects introduced by the experimental condition that were independent of the constructs under investigation. This covariation of relevant and irrelevant task complexity effects is why a "traditional" Brinley analysis often fails to find task-specific effects. Noting that "application of OLS regression to nested and unbalanced RT data produces regression coefficient estimates that are an uninterpretable mixture of between-experiment and within-experiment effects", Sliwinski and Hall (1998, p.165) proposed hierarchical linear models as an alternative approach to the meta-analysis of Brinley plot data. The use of hierarchical linear models was not applicable to the current data set, because it requires estimating slope as well as intercept parameters within each experiment, stable estimates of which in turn require that each factor in the experiment has at least three to four levels. The approach developed here to tackle the same problem—elimination of irrelevant between-experiment effects—is better suited to the type of data under investigation, which was generated by experimental designs consisting of orthogonal manipulation (guided by an explicit interaction hypothesis) of two or more factors, each of which had only a few levels. Admittedly, it might not always be easy to identify the appropriate regression model. Here, the regression was modeled after an explicit model of slowing in the tasks at hand, which might not always be available. On the other hand, the theoretical model I used was rather simplistic, and some sort of model is often, if not always, at least implicity guiding investigations. Making the model explicit helps to predict the expected pattern in the Brinley plot, which might take a form quite different from the one expected without an explicit model.

Process-specific Slowing

Inhibitory deficit? We have identified conditions under which the slope of the Brinley function is near one, and other conditions under which it is about 1.7–1.9. The latter slope was observed when the task posed working and/or episodic memory demands, while a slowing factor near one was observed for conditions with relatively small memory demands. The fact that the slope is near one is astonishing, because the line connected points from congruent and incongruent Stroop conditions, i.e., conditions that had previously been thought to be differentially age-sensitive. With the usual reserve, these results might be used to further specify the inhibitory deficit account of aging. There appears to be a specific age deficit related to ignoring irrelevant information, however, the level at which the information influences processing is critical. There are no age differences in overcoming pre-activation of a response if the association between a target concept and a response is well-established. Importantly, this is not limited to cases where the response is automatically activated: color

---

[82] In fact, compared to the 4 out of 19 studies that did use a single item format, the Stroop effects were larger with the blocked format.

[83] Similar learning effects were observed in the present experiments. These were not reported, because they did not affect the main results, which were still valid even if only trials from the last block were analyzed.

naming is not usually thought to be a highly automatic task. Instead, reading is much more automatic, and therefore, the wrong response is initially primed. The fact that there are no age differences in the compatibly mapped versions of the Stroop task is difficult to reconcile with the inhibitory deficit predictions that "very well-learned responses are particularly difficult to control, at least if they are wrong" and that "it is difficult to prevent strong responses from being produced, and this is particularly true for older adults." (Zacks & Hasher, 1997, p. P276).

Rather, what is important is whether or not the association of an action-relevant concept and a response is newly acquired. In conditions where responding was based on such "episodic" associations, old adults were much more susceptible to interference than young adults. Thus episodic retrieval and/or the maintenance of temporary associations appears to be particularly interference-prone in old age. This is at least partly compatible with the assumption that old adults have an inhibitory deficit that manifests itself in the "failure to suppress irrelevant retrieval pathways at the time of testing." (Zacks & Hasher, 1997, p. P275). However, the current results limit the scope of the inhibitory deficit framework to inhibition in short-term and episodic memory: episodic retrieval appears to be interference-prone in old age.

Spatial and verbal task domains. A question that cannot be answered decisively with the present data is whether the observed amplification is really caused by the proposed mechanism of a reduced reliability of episodic accumulators, or due to an age-differential strategic choice about the internal code used to represent the arbitrary mappings. The proposed explanation is that episodic task components lead to an amplification of early difficulty effects. The alternative explanation would hold that old adults prefer to code arbitrary rules verbally, while young adults can more flexibly choose between spatial and verbal coding strategies.

A spatial coding strategy might be particularly beneficial in tasks using manual responses, where verbal coding would lead to potentially large interference effects for old adults, because the verbal distractor—although it does not directly interfere with the response codes—interferes with the module used to represent and maintain the S-R mapping. Spatial coding would lead to smaller interference effects, because the verbal distractor does not interfere either with the response code or the code used for mapping maintenance. Thus, while the Stroop task with an arbitrary mapping might not strictly be a spatial task, a strategic choice to internally represent the mappings using spatial codes could be beneficial.

In tasks using a compatible mapping, there is no need to represent arbitrary rules. However, in most of the present experiments, compatible target-response mappings used a combination of vocal responses and phonological distractors. Because responses in vocal color-naming tasks rely on a the activation of a phonological code by an associated lexical entry, these might be considered verbal tasks.

The broad-domain specific general slowing theory proposed by Myerson and colleagues that assumes different slowing factors for verbal/lexical and spatial/nonlexical task domains might therefore be able to at least partly account for the present results. I still prefer the episodic reliability model, because the predictions were directly derived from this account and were largely fulfilled, whereas the spatial/verbal distinction needs to be stretched quite a bit to rather allusively explain the current results. Furthermore, Mayr and Kliegl (1993) and Verhaeghen et al. (2002) provide evidence that varying the degree of working memory demands can lead to two very different linear Brinley slopes even within the spatial task domain.

Nevertheless, more research is needed before the alternative explanation can be discounted. Experiments that could shed light on the true mechanism include the use of an articulatory suppression technique to prevent a verbal coding of arbitrary rules (e.g., Emerson & Miyake, 2003; Miyake, Emerson, Padilla, & Ahn, 2004), the variation of distractor type (e.g. color word vs. arrow vs. direction word) in a manual response Stroop task, or an investigation of spatial Stroop tasks with more response alternatives than in Experiment 5.

Manual vs. vocal responding. A spatial Stroop task with more response alternatives could also help to investigate another open question, namely, whether the lack of an age × complexity effect with compatible mappings is independent of response modality. In most of the current experiments, there was a sizeable interaction of Age and Response modality. Although this might have partly been caused (a) by the confounding of response modality with memory demands, and (b) by an age-differential speed accuracy trade-off in the manual modality, a similar interaction has been reported in the literature even in simple reaction time tasks (Nebes, 1978). It is at least a possibility that response modality might contribute to the age × complexity effect. If modality-specific Brinley slopes were found, this would clearly suggest a cognitive modality effect, because the peripheral demands of pressing a button do not change with, for example, the number of response alternatives. This would point at a locus of specific age effects in cognitive modules implementing response-related processes (such as premotor areas, (pre-)supplementary motor areas, and possibly ACC).

One result from the current meta-analysis that suggests some modality-specific slowing is the fact that the intercepts for the arbitrary-manual regression lines were consistently smaller than the intercepts for the arbitrary-vocal lines. In a standard analysis, using only two intercept and slope parameters, this could lead to somewhat smaller slope estimates for the vocal responses. However the current interpretation is that response modality does not change the slope, but only the intercept and does therefore not contribute to the complexity effect in the processes under investigation.

While the results seem to indicate that response modality does not change Brinley slope, the conclusion is not a very firm one, because we lack data points from one critical condition. In the current series of experiments, compatible mappings in the Stroop task were primarily realized using vocal responding. A manual/compatible condition was only tested in Experiment 5, where responding was generally fast because of the small number of response alternatives. Data from manual-compatible conditions with a higher overall difficulty are needed to evaluate whether the proposed age-equivalence in complexity effects with compatible mappings also holds for the manual response modality.

Episodic demands and related conceptions. The conception of episodic memory employed in the current thesis has a large degree of overlap with "controlled processing", as discussed in the literature on controlled and automatic processing (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). To vary task rule arbitrariness I mainly used stimulus-response compatibility manipulations. The continuum of stimulus-response compatibility that is discussed in the SRC literature is almost congruent with the continuum of automatic and controlled processes. Compatibility is a relative term, and highly compatible mappings are usually defined by a long learning history in ontogeny and sometimes even phylogeny. Responding with compatible S-R mappings is viewed as being governed by automatic processing, which is described as the stimulus-triggered retrieval from long-term memory that does not require executive or attentional capacity. On the other hand, the characteristic feature of an "episodic" task in the current view is that it relies on arbitrary task rules, and responding with arbitrary mappings is compared to a serial "table lookup" process. Arbitrary rules are by definition not overlearned, but freshly acquired within the experimental context. Hence, they require controlled processing, which is "a temporary activation of a sequence of elements that can be set up quickly and easily but requires attention, is capacity-limited (usually serial in nature), and is controlled" by the subject's homunculus (Schneider & Shiffrin, 1977, p.1).

Thus, the current results may also be regarded as further evidence for an age-related deficit in controlled processes, which has been documented by Rogers and colleagues (Fisk & Rogers, 1991; Rogers & Fisk, 1991). An increase in automaticity leads to a decreased mediation of episodic accumulators, which can be more and more bypassed, for example because a direct route (long-term "semantic" association) is established . Recent results (Scialfa, Jenkins, Hamaluk, & Skaloud, 2000) indicate that some aspects of the development of automaticity are age invariant, as long as arbitrary task aspects are kept at a minimum. Viewed from a different angle, a distinction between semantic memory processes and episodic or executive processes could be a critical mediator of age effects in the development of automaticity.

The current distinction between arbitrary and compatible mappings in reaction time tasks maps relatively well onto the episodic/semantic distinction in memory tasks (although the concept employed there is usually more elaborate). A significant decline in episodic recall is one of the most robust findings in cognitive aging (Allen, 1991; Craik & Jennings, 1992; Cerella, 1985). Similarly well-documented is the fact that semantic knowledge and crystallized intelligence is relatively unaffected by aging (e.g., Krampe & Ericsson, 1996; Allen, Sliwinski, Bowie, & Madden, 2002; Verhaeghen et al., 1997; Laver & Burke, 1993). A distinguished contribution of the current study is that episodic memory demands play a large role in determining the age × complexity effect even in reaction time tasks that pose apparently trivial demands on short-term memory (at least if compared to memory tasks proper).

Conclusions

What does a Brinley plot tell us about aging? Certainly, the result of a positive Brinley slope in most task domains must be accounted for. As Salthouse (1996) states, slowing does not need to be equivalent in magnitude for different cognitive tasks, but it is general in the sense that most of the variability in age-related slowing across many different tasks is common or shared. Although the present results are incompatible with general slowing in the sense of a basal, single-factor explanation in a serial information processing framework, I do not have a problem with this interpretation.

However, in my opinion, the typical Brinley plot pattern tells us more about task complexity than about "general" age-related slowing. It might possibly suggest something about the way that task complexity (young RT) covaries with the episodic working memory demands of the task. The present series of experiments can be regarded as an attempt at decomposing reaction time effects into working memory-related and -unrelated effects. It was shown that Brinley slopes were rather close to one for the latter. Importantly, the (Stroop) effect under investigation is typically considered "cognitive", and sometimes even considered a prime indicator for executive processing. Hence, the current result of a Brinley slope near one cannot be interpreted to merely indicate age equivalence in peripheral processes.

What kind of processing could lead to the observed covariations of working memory demands, task complexity, and age effects? The Episodic Accumulator Model assumes that episodic information is unreliably represented, particularly so in old age, and that leakage has to be compensated for by incoming information before the response threshold is reached. The idea that information loss is compensated for by a larger number of processing steps is also found in the model proposed by Mayr and Kliegl (1993) to explain age effects in working memory. These authors note that the greater number of processing steps required means that (under coordinative complexity conditions) young and old adults differ in the algorithms used, thereby violating a central assumption of extant slowing models.

Here is a classic description of this "correspondence axiom" (or homogeneity assumption) linking young and old adults' algorithms (Cerella, 1990, p.215): "Young and old adults are assumed to be performing the same computation, which is to say that age operates solely on the integrity and not the logic of a network." Yet it appears that what consitutes "the logic" depends on the level of description. At a macroscopic level of description, the algorithm may read "perform a lookup of value A, given key B", and may be equivalent between age groups. At a more microscopic level however, the same algorithm might involve a loop until a test value reaches a criterion. The criterion might be reached later for old adults, because the rate of information accumulation for the test value within the loop is smaller for old adults—for example, because the accumulator is leaky.[84]

Is circular, loop-like processing with information loss a reasonable assumption? The Episodic Accumulator Model does not make specific assumptions about short-term memory or episodic retrieval itself. However, many contemporary neuro-computational models of short-term memory are available to fill the gap (e.g., Wang, 2001; O'Reilly, Braver, & Cohen, 1999; Amit, 1995; Durstewitz, Kelc, & Güntürkün, 1999; Deco & Rolls, 2003; Lisman, Fellous, & Wang, 1998). Most of these models are variants of the Hopfield network which assume that there is Hebbian-type recurrent activation, and a leaky integrator is often chosen as the single neuron processing unit. A larger leakage term in old age appears to be well-suited to incorporate the current ideas of a reduced reliability of episodic accumulators. A core feature of the strong processing assumptions made by these models is the circulation of activation. With recurrent processing, performing the same computation does not necessarily involve the same number of processing steps. Instead, the number of steps depends on the value of the leakage term. In other words, if more information is lost during reverberation, then information will have to circulate for longer until a stable attractor is reached.

With respect to aging, although the same macroscopic-level algorithm is applied, more microscopic processing steps will have to be performed by old adults. Thus at this level of analysis, the conception of the correspondence axiom has to be refined. If one is willing to give up the strong form of the correspondence axiom, then circulation of activation might therefore be one possibility to rejoin a basal deficit cognitive aging theory with the current findings as well as with the over- and underestimations reported in the Introduction.

More generally, contemporary views of cognition regard information processing in the "perception-action cycle" as determined by hierarchically organized feedback loops at multiple levels (Fuster, 2004; Koechlin, Ody, & Kouneiher, 2003). The translation of environmental stimuli into goal-directed actions is regulated by external and several levels of internal feedback. Only relatively low levels of internal control are involved with automatic and well-rehearsed responses to simple stimuli. As stimuli or responses become more complex, higher-level control modules become active additionally. Yet higher levels of control are needed in addition when the stimulus-response associations are not well-practiced, when the task relies on short-term (in particular, multi-modal) associations and when the task rules themselves have to be retrieved, depending on the context. According to this conception, more executive and episodic tasks require feedback loops at increasingly high levels of prefrontal cortex, thereby increasing the average number of activation circulations until the criterion is reached. From this point of view, the current results can be taken to indicate that age effects are the more severe, the higher the level of control is called for.

In conclusion, the present research adds to the evidence suggesting that executive and episodic processes of working memory maintenance and retrieval are a source of a specific age-related deficit in cognition. While old adults appear to have a specific deficit at this level of analysis, this does not completely rule out that a more general deficit is responsible at a more microscopic level of description. A possible scenario that can save basal-deficit general slowing theories requires the assumption that old adults need more elementary processing steps to achieve the same computation. The present results therefore aid in specifying the type of computation underlying age-related slowing as well as task complexity: A tentative hypothesis is that episodic task demands positively covary with the amount of circulatory processing required to perform a task.

References

Allen, P. A. (1991). On age differences in processing variability and scanning speed. *Journals of Gerontology*, *46*(5), P191–P201.

Allen, P. A., Sliwinski, M., Bowie, T., & Madden, D. J. (2002). Differential age effects in semantic and episodic memory. *Journals of Gerontology (B)*, *57B*(2), P173–P186.

Allport, A., & Wylie, G. (2000). Task-switching: Positive and negative priming of task-set. In G. W. Humphreys, J. Duncan, et al. (Eds.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). London: Oxford University Press.

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance 15: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Alluisi, E. A., Strain, G. S., & Thurmond, J. B. (1964). Stimulus-response compatibility and the rate of gain of information. *Psychonomic Science*, *1*, 111–112.

---

[84] At a yet more microscopic level, retrieval (i.e., the evidence accumulation processes leading to criterion) might be determined by the "degree of resonance" that a probe evokes in a memory set. In a prominent model, "evidence is accumulated in parallel from each probe-memory item comparison, and each comparison is modeled by a continuous random walk process" (Ratcliff, 1978, p.59). At this level of description, the possibility of confusion between elements of the memory set (or the likelihood of intrusion errors) appears to be a major determinant of item-differential drift rates. Likelihood of confusion is enhanced by a degradation of the memory set representation. This might explain why I observed age-enhanced interference at episodic retrieval as well as why over-proportional age effects have been reported for task-switching situations with ambiguous stimuli and full response set overlap (Mayr, 2001) and why old adults are highly susceptible to intrusion errors in working memory (Oberauer, 2001).

Amit, D. J. (1995). The hebbian paradigm reintegrated: Local reverberations as internal representations. *Behavioral & Brain Sciences*, *18*(4), 617–657.

Asaad, W. F., Rainer, G., & Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron*, *21*, 1399–1407.

Babcock, R. (1994). Analysis of adult age differences on the Raven's Advanced Progressive Matrices test. *Psychology & Aging*, *9*(2), 303–314.

Baddeley, A. (1986). *Working memory*. New York: Clarendon Press/Oxford University Press.

Baddeley, A. (1992a). Is working memory working? The fifteenth Bartlett lecture. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *44A*(1), 1–31.

Baddeley, A. (1992b). Working memory. *Science*, *255*(5044), 556–559.

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

Baddeley, A. (2001). Is working memory still working? *American Psychologist*, *56*(11), 849–864.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–90). New York: Academic Press.

Baeckman, L., & Nilsson, L.-G. (1996). Semantic memory functioning across the adult life span. *European Psychologist*, *1*(1), 27–33.

Barone, P., & Joseph, J. (1989). Role of the dorsolateral prefrontal cortex in organizing visually guided behavior. *Brain, Behavior and Evolution*, *33*(2-3), 132–135.

Bashore, T., Osman, A., & Heffley, E. (1989). Mental slowing in elderly persons: a cognitive psychophysiological analysis. *Psychology & Aging*, *4*(2), 235–244.

Bichot, N. P., Schall, J. D., & Thompson, K. G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, *381*(6584), 697–699.

Biederman, I., & Kaplan, R. (1970). Stimulus discriminability and S-R compatibility: Evidence for independent effects in choice reaction time. *Journal of Experimental Psychology*, *86*(3), 434–439.

Birren, J. (1959). Principles of research on aging. In J. Birren (Ed.), *Handbook of aging and the individual.* (pp. 3–42). Chicago, IL: Univ. of Chicago Press.

Birren, J. E., & Botwinick, J. (1955a). Age differences in finger, jaw, and foot reaction time to auditory stimuli. *Journal of Gerontology*, *10*, 430–432.

Birren, J. E., & Botwinick, J. (1955b). Speed of response as a function of perceptual difficulty and age. *Journal of Gerontology*, *10*, 433–436.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*(6758), 179–181.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., Janowsky, J. S., Taylor, S. F., Yesavage, J. A., Mumenthaler, M. S., Jagust, W. J., & Reed, B. R. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, *130*(4), 746–763.

Brink, J. M., & McDowd, J. M. (1999). Aging and selective attention: An issue of complexity or multiple mechanisms? *Journals of Gerontology (B)*, *54B*(1), P30–P33.

Brinley, J. F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In A. Welford & J. Birren (Eds.), *Behavior, aging and the nervous system* (pp. 114–149). Springfield, IL: Thomas.

Burke, D. M. (1997). Language, aging, and inhibitory deficits: Evaluation of a theory. *Journals of Gerontology (B)*, *52B*(3), P254–P264.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222.

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.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747–749.

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, *98*(1), 67–83.

Cerella, J. (1990). Aging and information-processing rate. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 201–221). San Diego, CA: Academic Press, Inc.

Cerella, J. (1994). Generalized slowing in Brinley plots. *Journals of Gerontology*, *49*(2), P65–P71.

Cerella, J., Poon, L., & Williams, D. (1980). Age and the complexity hypothesis. In L. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 332–340). Washington, DC: American Psychological Association.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficits. *Psychological Bulletin*, *79*(6), 380–385.

Charness, N., & Campbell, J. I. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General*, *117*(2), 115–129.

Clark, H. (1973). Space, time semantics, and the child. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (p. 27-63). New York: Academic Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance 15: Conscious and nonconscious information processing.* (Vol. XV, pp. 453–476). Cambridge, MA: MIT Press.

Cohn, N. B., Dustman, R. E., & Bradford, D. C. (1984). Age-related decrements in Stroop color test performance. *Journal of Clinical Psychology*, *40*(5), 1244–1250.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256.

Comalli, P. E. J., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology*, *100*, 47–53.

Constantinidis, C., & Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology*, *76*(2), 1352–1355.

Conway, A. R., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, *4*(6), 577–591.

Cornelius, S. W. (1984). Classic pattern of intellectual aging: Test familiarity, difficulty, and performance. *Journal of Gerontology*, *39*(2), 201–206.

Cowan, N. (1995). *Attention and memory: An integrated framework.* London: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York: Cambridge University Press.

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition.* (pp. 51–110). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Crowder, R. G. (1967). Short-term memory for words with a perceptual-motor interpolated activity. *Journal of Verbal Learning & Verbal Behavior*, *6*(5), 753–761.

Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the Stroop color-word task. *Perception & Psychophysics*, *11*(4), 274–276.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*(4), 450–466.

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

Deco, G., & Rolls, E. T. (2003). Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *European Journal of Neuroscience*, *18*(8), 2374–2390.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: Dissociable causes of interference effects in conflict situations. *Acta Psychologica*, *101*(2-3), 379–394.

De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception & Performance*, *20*(4), 731–750.

DeSoto, M. C., Fabiani, M., Geary, D. C., & Gratton, G. (2001). When in doubt, do it both ways: Brain evidence of the simultaneous activation of conflicting motor responses in a spatial Stroop task. *Journal of Cognitive Neuroscience*, *4*(4), 523–537.

Doose, G., & Feyereisen, P. (2001). Task specificity in age-related slowing: Word production versus conceptual comparison. *Journals of Gerontology (B)*, *56B*(2), P85–P87.

Dulaney, C. L., & Rogers, W. A. (1994). Mechanisms underlying reduction in Stroop interference with practice for young and old adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*(2), 470–484.

Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception & Performance*, *10*(5), 622–639.

Duncan, J. (1977). Response selection in spatial choice reaction time tasks. In S. Dornic (Ed.), *Attention and performance VI* (pp. 49–61). Hillsdale, NJ: Erlbaum.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*(10), 475–483.

Duncan-Johnson, C. C., & Kopell, B. S. (1980). The locus of interference in a Stroop task - when you read "blue", do you see "red"? *Psychophysiology*, *17*(3), 308–309.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of interference. *Science*, *214*(4523), 938–940.

Dunn, J. C., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*(1), 91–101.

Durgin, F. H. (2000). The reverse Stroop effect. *Psychonomic Bulletin & Review*, *7*(1), 121-125.

Durstewitz, D., Kelc, M., & Güntürkün, O. (1999). A neurocomputational theory of the dopaminergic modulation of working memory functions. *Journal of Neuroscience*, *19*(7), 2807–2822.

Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, *1*(2), 106–120.

Egeth, H. (1977). Attention and preattention. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 277–320). New York: Academic Press.

Eimer, M. (1995). Stimulus-response compatibility and automatic response activation: Evidence from psychophysiological studies. *Journal of Experimental Psychology: Human Perception & Performance*, *21*(4), 837–854.

Eimer, M. (1997). The Lateralized Readiness Potential as an on-line measure of automatic response activation in S-R compatibility situations. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus-response compatibility* (Vol. 118, pp. 51–73). Amsterdam: Elsevier.

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, 148–168.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–24.

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

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, *125*(6), 777–799.

Felbrich, A. (2000, January). *Age effects in visual search.* Diploma Thesis Reg. No. 126962. Department of Cognitive Psychology. University of Potsdam.

Fisk, A. D., & Fisher, D. L. (1994). Brinley plots and theories of aging: The explicit, muddled, and implicit debates. *Journals of Gerontology*, *49*(2), P81–P89.

Fisk, A. D., & Rogers, W. A. (1991). Toward an understanding of age-related memory and visual search effects. *Journal of Experimental Psychology: General*, *120*(2), 131–149.

Fisk, A. D., Rogers, W. A., & Giambra, L. M. (1990). Consistent and varied memory/visual search: Is there an interaction between age and response-set effects? *Journals of Gerontology*, *45*(3), P81–P87.

Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: correspondence among paired elements within stimulus and response code. *Journal of Experimental Psychology*, *48*, 483–492.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*(1953), 199–210.

Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and encoding factors in ignoring irrelevant information. *Memory & Cognition*, *7*(2), 86–94.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, *18*(4), 1030–1044.

Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychologica*, *30*, 96–103.

Fuster, J. M. (2004). Upper processing stages of the perceptionaction cycle. *Trends in Cognitive Sciences*, *8*(4), 143–145.

Geary, D. C., Frensch, P. A., & Wiley, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in younger and older adults. *Psychology & Aging*, *8*(2), 242–256.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, *44*(3), 297–337.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception & Performance*, *8*(6), 875–894.

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception & Performance*, *10*(5), 640–654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*(1), 13–42.

Gopher, D., & Navon, D. (1980). How is performance limited: Testing the notion of central capacity. *Acta Psychologica*, *46*(3), 161–180.

Gottlob, L. R., & Madden, D. J. (1998). Time course of allocation of visual attention after equating for sensory differences: An age-related perspective. *Psychology & Aging*, *13*(1), 138–149.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506.

Greenwood, P. M., & Parasuraman, R. (1994). Attentional disengagement deficit in nondemented elderly over 75 years of age. *Aging, Neuropsychology, & Cognition*, *1*(3), 188–202.

Greenwood, P. M., Parasuraman, R., & Haxby, J. V. (1993). Changes in visuospatial attention over the adult lifespan. *Neuropsychologia*, *31*(5), 471–485.

Gurney, K., Prescott, T., & Redgrave, P. (2001a). A computational model of action selection in the basal ganglia. II. Analysis and simulation of behavior. *Biological Cybernetics*, *84*, 411–423.

Gurney, K., Prescott, T., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biological Cybernetics*, *84*, 401–410.

Hale, S., & Myerson, J. (1996). Experimental evidence for differential slowing in the lexical and nonlexical domains. *Aging, Neuropsychology, & Cognition*, *3*(2), 154–165.

Hale, S., Myerson, J., Faust, M., & Fristoe, N. (1995). Converging evidence for domain-specific slowing from multiple nonlexical tasks and multiple analytic methods. *Journals of Gerontology (B)*, *50B*(4), P202–P211.

Hale, S., Myerson, J., & Wagstaff, D. (1987). General slowing of nonverbal information processing: Evidence for a power law. *Journal of Gerontology*, *42*(2), 131–136.

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

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427–430.

Hartley, A. A. (1993). Evidence for the selective preservation of spatial selective attention in old age. *Psychology & Aging*, *8*(3), 371–379.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*(1), 163–169.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.

Hertzog, C., & Rypma, B. (1991). Age differences in components of mental-rotation task performance. *Bulletin of the Psychonomic Society*, *29*(3), 209–212.

Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, *83*(2, Pt. 1), 299–303.

Hommel, B. (1998). Automatic stimulus-response translation in dual-task performance. *Journal of Experimental Psychology: Human Perception & Performance*, *24*(5), 1368–1384.

Hommel, B., & Eglau, B. (2002). Control of stimulus-response translation in dual-task performance. *Psychological Research/Psychologische Forschung*, *66*(4), 260–273.

Humphrey, D. G., & Kramer, A. F. (1997). Age differences in visual search for feature, conjunction, and triple-conjunction targets. *Psychology & Aging*, *12*(4), 704–717.

Jenkins, L., Myerson, J., Joerding, J. A., & Hale, S. (2000). Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. *Psychology & Aging*, *15*(1), 157–175.

Jensen, A. R. (1998). *Human evolution, behavior, and intelligence*. Westport, CT: Praeger Publishers/Greenwood Publishing.

Jensen, A. R. (2000). IQ tests, psychometric and chronometric g, and achievement. Reply to Kush on Jensen on intelligence-g-factor. [11 paragraphs]. *Psycoloquy [Online serial]*, *11*(14 (29)).

Jonides, J., Badre, D., Curtis, C., Thompson-Schill, S. L., & Smith, E. E. (2002). Mechanisms of conflict resolution in prefrontal cortex. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 233–245). London: Oxford University Press.

Jordan, T. C., & Rabbitt, P. M. A. (1977). Response times to stimuli of increasing complexity as a function of aging. *British Journal of Psychology*, *68*(2), 189–201.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*(2), 137–172.

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

Kahneman, D., & Chajczyk, D. (1983). Tests of the automaticity of reading: Dilution of Stroop effects by color-irrelevant stimuli. *Journal of Experimental Psychology: Human Perception & Performance*, *9*(4), 497–509.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169-183.

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, & Cognition*, *26*(2), 336–358.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70.

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

Keele, S. W. (1969). Repetition effect: A memory dependent process. *Journal of Experimental Psychology*, *80*, 243–248.

Keele, S. W. (1972). Attention demands of memory retrieval. *Journal of Experimental Psychology*, *93*(2), 245–248.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General*, *129*(3), 340–360.

Kirasic, K. C. (1991). Spatial cognition and behavior in young and elderly adults: Implications for learning new environments. *Psychology & Aging*, *6*(1), 10–18.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology*, *77*(4), 576–588.

Kliegl, R., & Lindenberger, U. (1993). Modeling intrusions and correct recall in episodic memory: Adult age differences in encoding of list context. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*(3), 617–637.

Kliegl, R., Mayr, U., & Krampe, R. T. (1994). Time-accuracy functions for determining process and person differences: An application to cognitive aging. *Cognitive Psychology*, *26*(2), 134–164.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181–1185.

Kornblum, S. (1973). Sequential effects in choice reaction time: A tutorial review. In S. Kornblum (Ed.), *Attention and performance IV*. New York: Academic Press.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility: A model and taxonomy. *Psychological Review*, *97*(2), 253–270.

Kramer, A. F., Hahn, S., Cohen, N. J., Banich, M. T., McAuley, E., Harrison, C. R., Chason, J., Vakil, E., Bardell, L., Boileau, R. A., & Colcombe, A. (1999). Ageing, fitness and neurocognitive function. *Nature*, *400*(6743), 418–419.

Kramer, A. F., & Larish, J. L. (1996). Aging and dual-task performance. In W. A. Rogers, A. D. Fisk, et al. (Eds.), *Aging and skilled performance: Advances in theory and applications* (pp. 83–112). Hillsdale, NJ: Lawrence Erlbaum Associates.

Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General*, *125*(4), 331–359.

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

La Heij, W., van der Heijden, A. H., & Schreuder, R. (1985). Semantic priming and Stroop-like interference in word-naming tasks. *Journal of Experimental Psychology: Human Perception & Performance*, *11*(1), 62–80.

Lammertyn, J., Fias, W., & Lauwereyns, J. (2002). Semantic influences on feature-based attention due to overlap of neural circuits. *Cortex*, *38*(5), 878–882.

Laver, G. D., & Burke, D. M. (1993). Why do semantic priming effects increase in old age? A meta-analysis. *Psychology & Aging*, *8*(1), 34–43.

Lawrence, B., Myerson, J., & Hale, S. (1998). Differential decline of verbal and visuospatial processing speed across the adult life span. *Aging, Neuropsychology, & Cognition*, *5*(2), 129–146.

Lehrl, S. (1995). *Mehrfachwahl-Wortschatz-Test (Form B)* [Multiple-choice vocabulary intelligence test (Form B/MSVT-B)] (3rd ed.). Erlangen: Straube. (Original work published 1977)

Lehrl, S., Daun, H., & Schmidt, R. (1971). Mehrfachwahl-Wortschatz-Test (HAWIE). Eine Abwandlung des HAWIE-Wortschatztests als Kurztest zur Messung der Intelligenz Erwachsener [Multiple-choice vocabulary intelligence test (WAIS)]. *Archiv fuer Psychiatrie und Nervenkrankheiten*, *214*, 353–364.

Lehrl, S., & Fischer, B. (1988). The basic parameters of human information processing: Their role in the determination of intelligence. *Personality & Individual Differences*, *9*(5), 883–896.

Levinson, S. C. (2003). Spatial language. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol. 4, pp. 131–137). London: Nature Publishing Group.

Li, S.-C., Lindenberger, U., & Frensch, P. A. (2000). Unifying cognitive aging: From neuromodulation to representation to cognition. *Neurocomputing*, *32-33*, 879–890.

Li, S.-C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*, *5*(11), 479–486.

Lien, M.-C., & Proctor, R. W. (2002). Stimulus-response compatibility and psychological refractory period effects: Implications for response selection. *Psychonomic Bulletin & Review*, *9*(2), 212–238.

Light, L. L., LaVoie, D., Valencia-Laver, D., Owens, S. A., et al. (1992). Direct and indirect measures of memory for modality in young and older adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*(6), 1284–1297.

Lima, S. D., Hale, S., & Myerson, J. (1991). How general is general slowing? Evidence from the lexical domain. *Psychology & Aging*, *6*(3), 416–425.

Lindenberger, U., Mayr, U., & Kliegl, R. (1993). Speed and intelligence in old age. *Psychology & Aging*, *8*(2), 207–220.

Lisman, J. E., Fellous, J.-M., & Wang, X.-J. (1998). A role for NMDA-receptor channels in working memory. *Nature Neuroscience*, *1*(4), 273–275.

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

Logan, G. D. (1980). Short-term memory demands of reaction-time tasks that differ in complexity. *Journal of Experimental Psychology: Human Perception & Performance*, *6*(2), 375–389.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527.

Logan, G. D., & Gordon, R. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*(2), 393–434.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition*, *7*(3), 166–174.

Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses. *Journal of Experimental Psychology: Human Perception & Performance*, *24*(3), 978–992.

Logan, G. D., Zbrodoff, N. J., & Williamson, J. (1984). Strategies in the color-word Stroop task. *Bulletin of the Psychonomic Society*, *22*(2), 135–138.

Lohman, D. (1988). Spatial abilities as traits, processes, and knowledge. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence (Vol. 4)* (pp. 181–248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lu, C.-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*(2), 174–207.

Lubbe, R. H. J. Van der, & Verleger, R. (2002). Aging and the Simon task. *Psychophysiology*, *39*(1), 100–110.

Luo, C. R. (1999). Semantic competition as the basis of Stroop interference: Evidence from color-word matching tasks. *Psychological Science*, *10*(1), 35–40.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199-207.

MacDonald III, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*(5472), 1835-1838.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203.

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, *26*(2), 201–211.

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*(1), 126–135.

MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences*, *4*, 383–391.

Madden, D. J. (1986). Adult age differences in the attentional capacity demands of visual search. *Cognitive Development*, *1*(4), 335–363.

Madden, D. J., Pierce, T. W., & Allen, P. A. (1993). Age-related slowing and the time course of semantic priming in visual word identification. *Psychology & Aging*, *8*(4), 490–507.

Martínez, A., Anllo-Vento, L., Sereno, M., Frank, L., Buxton, R., Dubowitz, D., Wong, E., Hinrichs, H., Heinze, H., & Hillyard, S. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*(4), 364–369.

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology & Aging*, *16*(1), 96–109.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*(5), 450–452.

Mayr, U., & Kliegl, R. (1993). Sequential and coordinative complexity: Age-based processing limitations in figural transformations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*(6), 1297–1320.

Mayr, U., & Kliegl, R. (2000). Complex semantic processing in old age: Does it stay or does it go? *Psychology & Aging*, *15*(1), 29–43.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*(3), 362–372.

Mayr, U., Kliegl, R., & Krampe, R. T. (1996). Sequential and coordinative processing dynamics in figural transformations across the life span. *Cognition*, *59*(1), 61–90.

Mayr, U., & Laubrock, J. (2002). *DFG-Projektbericht: Episodische Akkumulatoren und kognitive Verlangsamung im Altern (MA 2154/1-1).*

Mayr, U., & Liebscher, T. (2001). Is there an age deficit in the selection of mental sets? *European Journal of Cognitive Psychology*, *13*(1/2), 47–70.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*(4477), 77–80.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287–330.

McDowd, J. M. (1997). Inhibition in attention and aging. *Journals of Gerontology (B)*, *52B*(6), P265–P273.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*(3), 211–253.

Meiran, N., Gotler, A., & Perlman, A. (2001). Old age is associated with a pattern of relatively intact and relatively impaired task-set switching abilities. *Journals of Gerontology (B)*, *56B*(2), P88–P102.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–203.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*(16), 5154–5167.

Miller, E. M. (1994). Intelligence and brain myelination: A hypothesis. *Personality & Individual Differences*, *17*(6), 803–832.

Miller, J., & Anbar, R. (1981). Expectancy and frequency effects on perceptual and motor systems in choice reaction time. *Memory & Cognition*, *9*(6), 631–641.

Miller, J., & Hackley, S. A. (1992). Electrophysiological evidence for temporal overlap among contingent mental processes. *Journal of Experimental Psychology: General*, *121*(2), 195–209.

Miyake, A., Emerson, M. J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression in the random task cuing paradigm. *Acta Psychologica*, *115*, 123–142.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, *27*, 84–92.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178.

Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental Psychology*, *25*(3), 387–397.

Mueller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception & Performance*, *15*(2), 315–330.

Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review*, *97*(4), 475–487.

Myerson, J., Wagstaff, D., & Hale, S. (1994). Brinley plots, explained variance, and the analysis of age differences in response latencies. *Journals of Gerontology*, *49*(2), P72–P80.

Navon, D. (1984). Resources–a theoretical soup stone? *Psychological Review*, *91*(2), 216–234.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*(3), 214–255.

Nebes, R. D. (1978). Vocal versus manual response as a determinant of age difference in simple reaction time. *Journal of Gerontology*, *33*(6), 884-889.

Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*(6), 1291–1305.

Neill, W. T., & Valdes, L. A. (1992). Persistence of negative priming: Steady state or decay? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*(3), 565–576.

Neumann, O. (1980). *Informations-Selektion und Handlungsteuerung [selection of information and control of action]*. Unpublished doctoral dissertation, University of Bochum.

Neumann, O., & Klotz, W. (1994). Motor responses to nonreportable, masked stimuli: Where is the limit of direct parameter specification? In C. Umilta & M. Moscovitch (Eds.), *Attention and performance 15: Conscious and nonconscious information processing* (pp. 123–150). Cambridge, MA: MIT Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Newell, A., & Simon, H. A. (1963). GPS, a program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 279–293). New York: McGraw-Hill. (Reprinted from *Lernende Automaten*, pp. 109–124, by H. Billing, Ed., 1961, R. Oldenbourg)

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). New York: Plenum.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *75*(1), 44–64.

Nyberg, L., Forkstam, C., Petersson, K. M., Cabeza, R., & Ingvar, M. (2002). Brain imaging of human memory systems: Between-systems similarities and within-system differences. *Cognitive Brain Research*, *13*(2), 281–292.

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*(4), 948–957.

Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology*, *13*(1/2), 187–216.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167–193.

Oberauer, K., Wendland, M., & Kliegl, R. (2003). Age differences in working memory–the roles of storage and selective access. *Memory & Cognition*, *31*(4), 563–569.

Obler, L. K., & Albert, M. L. (1985). Language skills across adulthood. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 463–473). New York: Van Nostrand Reinhold.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*(11), 1203–1209.

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.

Pashler, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Pavese, A., & Umilta, C. (1998). Symbolic distance between numerosity and identity modulates Stroop interference. *Journal of Experimental Psychology: Human Perception & Performance*, *24*(5), 1535–1545.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Perfect, T. J. (1994). What can Brinley plots tell us about cognitive aging? *Journals of Gerontology*, *49*(2), P60–P64.

Peterson, B. S., Skudlarski, P., Gatenby, J. C., Zhang, H., Anderson, A. W., & Gore, J. C. (1999). An fMRI study of Stroop word-color interference: Evidence for cingulate subregions subserving multiple distributed attentional systems. *Biological Psychiatry*, *45*(10), 1237–1258.

Petrides, M. (1982). Motor conditional associative-learning after selective prefrontal lesions in the monkey. *Behavioural Brain Research*, *5*(4), 407–413.

Petrides, M. (1985). Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia*, *23*(5), 601–614.

Phaf, R. H., van der Heijden, A. H., & Hudson, P. T. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, *22*(3), 273–341.

Picard, N., & Strick, P. (2001). Imaging the premotor areas. *Current Opinion in Neurobiology*, *11*(6), 663–672.

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex*, *6*(3), 342–353.

Posner, M., & Snyder, C. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M. I., & Raichle, M. E. (1994). *Images of mind.* New York: Freeman/Scientific American Library.

Proctor, R. W. (1978). Sources of color-word interference in the Stroop color-naming task. *Perception & Psychophysics*, *23*(5), 413–419.

Proctor, R. W., & Reeve, T. G. (1990). Research on stimulus-response compatibility: Toward a comprehensive account. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus-response compatibility: An integrated perspective* (Vol. 65, pp. 483–494). Oxford: North-Holland.

Quintana, J., & Fuster, J. M. (1992). Mnemonic and predictive functions of cortical neurons in a memory task. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*, *3*(8), 721–724.

Rabbitt, P. M. A. (1967). Signal discriminability, S-R compatibility and choice reaction time. *Psychonomics Science*, *7*, 419–420.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review*, *7*(1), 1–25.

Redding, G. M., & Gerjets, D. A. (1977). Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual & Motor Skills*, *45*(1), 11–17.

Reed, T. E., & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence*, *16*(3-4), 259–272.

Riby, L., Perfect, T., & Stollery, B. (2004). Evidence for disproportionate dual-task costs in older adults for episodic but not semantic memory. *The Quarterly Journal of Experimental Psychology A*, *57*(2), 241–267.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*(1), 88–125.

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

Rogers, W. A., & Fisk, A. D. (1991). Are age differences in consistent-mapping visual search due to feature learning or attention training? *Psychology & Aging*, *6*(4), 542–550.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, *46*(1), 247–253.

Salthouse, T. A. (1978). The role of memory in the age decline in digit-symbol substitution performance. *Journal of Gerontology*, *33*(2), 232–238.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 400–426). New York: Van Nostrand Reinhold.

Salthouse, T. A. (1987). The role of representations in age differences in analogical reasoning. *Psychology & Aging*, *2*(4), 357–362.

Salthouse, T. A. (1988). Resource-reduction interpretations of cognitive aging. *Developmental Review*, *8*(3), 238–272.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Salthouse, T. A. (1992). What do adult age differences in the Digit Symbol Substitution test reflect? *Journals of Gerontology*, *47*(3), P121–P128.

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*(4), 535–543.

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

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *Journals of Gerontology (B)*, *50B*(6), P297–P306.

Salthouse, T. A., & Somberg, B. L. (1982). Isolating the age deficit in speeded performance. *Journal of Gerontology*, *37*(1), 59–63.

Sanders, A. F. (1995). Discrete vs. continuous processing: The fate of an incompletely processed perceptual dimension. *Acta Psychologica*, *90*(1-3), 211–227.

Schaie, K. W., & Parham, I. A. (1977). Cohort-sequential analyses of adult intellectual development. *Developmental Psychology*, *13*(6), 649–653.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Schwarz, W. (2003). Stochastic cascade processes as a model of multi-stage concurrent information processing. *Acta Psychologica*, *113*(3), 231–261.

Scialfa, C., Jenkins, L., Hamaluk, E., & Skaloud, P. (2000). Aging and the development of automaticity in conjunction search. *Journals of Gerontology (B)*, *55*(1), P27–46.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27.

Shallice, T., & Burgess, P. (1993). Supervisory control of action and thought selection. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 171–187). New York: Clarendon Press/Oxford University Press.

Sharma, D., & McKenna, F. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition*, *26*(5), 1033–1040.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190.

Shor, R. E. (1970). The processing of conceptual information on spatial directions from pictorial and linguistic symbols. *Acta Psychologica*, *32*(4), 346–365.

Shwartz, S. P., Pomerantz, J. R., & Egeth, H. E. (1977). State and process limitations in information processing: An additive factors analysis. *Journal of Experimental Psychology: Human Perception & Performance*, *3*(3), 402–410.

Sliwinski, M. J., & Hall, C. B. (1998). Constraints on general slowing: A meta-analysis using hierarchical linear models with random coefficients. *Psychology & Aging*, *13*(1), 164–175.

Smulders, F. T. Y., Kenemans, J. L., Schmidt, W. F., & Kok, A. (1999). Effects of task complexity in young and old adults: Reaction time and P300 latency are not always dissociated. *Psychophysiology*, *36*(1), 118–125.

Somers, D. C., Dale, A. M., Seiffert, A. E., & Tothell, R. B. H. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proceedings of the National Academy of the Sciences, USA*, *96*(4), 1663–1668.

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception & Performance*, *22*(2), 461–479.

Stafford, T. (2003). *Integrating psychological and neuroscientific constraints in models of Stroop processing and action selection*. Unpublished doctoral dissertation, Department of Psychology, University of Sheffield.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.

Sternberg, S. (1998). Inferring mental operations from reaction-time data: How we compare objects. In D. Scarborough & S. Sternberg (Eds.), *Methods, models, and conceptual issues: An invitation to cognitive science* (Vol. 4, pp. 365–454). Cambridge, MA: MIT Press.

Strayer, D. L., & Grison, S. (1999). Negative identity priming is contingent on stimulus repetition. *Journal of Experimental Psychology: Human Perception & Performance*, *25*(1), 24–38.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception & Performance*, *20*(3), 647–675.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability–and a little bit more. *Intelligence*, *30*(3), 261–288.

Teichner, W. H., & Krebs, M. J. (1974). Laws of visual choice reaction time. *Psychological Review*, *81*(1), 75–98.

Tewes, U. (1994). *HAWIE-R. Hamburg-Wechsler Intelligenztest für Erwachsene - Revision 1991* [Wechsler Adult Intelligence Scale - Revised (WAIS-R; Wechsler, D., 1981) - revised German version (1991)] (2nd ed.). Bern: Huber. (Original work published 1991)

Thomas, J. C., Waugh, N. C., & Fozard, J. L. (1978). Age and familiarity in memory scanning. *Journal of Gerontology*, *33*(4), 528-533.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579.

Treue, S., & Maunsell, J. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, *382*(6591), 539–541.

Tsang, P. S. (1998). Age, attention, expertise, and time-sharing performance. *Psychology & Aging*, *13*(2), 323–347.

Tun, P. A., Wingfield, A., & Stine, E. A. (1991). Speech-processing capacity in young and older adults: A dual-task study. *Psychology & Aging*, *6*(1), 3–9.

Turken, A. U., & Swick, D. (1999). Response selection in the human anterior cingulate cortex. *Nature Neuroscience*, *2*(10), 920–924.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, *28*(2), 127–154.

Ulrich, R., & Giray, M. (1989). Time resolution of clocks: Effects on reaction time measurement – good news for bad clocks. *British Journal of Mathematical and Statistical Psychology*, *42*, 1–12.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Verhaeghen, P., Cerella, J., Semenec, S., Leo, M., Bopp, K., & Steitz, D. (2002). Cognitive efficiency modes in old age: performance on sequential and coordinative verbal and visuospatial tasks. *Psychology & Aging*, *17*(4), 558–570.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology & Aging*, *13*(1), 120–126.

Verhaeghen, P., Kliegl, R., & Mayr, U. (1997). Sequential and coordinative complexity in time-accuracy functions for mental arithmetic. *Psychology & Aging*, *12*(4), 555–564.

Verhaeghen, P., Steitz, D., Sliwinski, M., & Cerella, J. (2003). Aging and dual-task performance: a meta-analysis. *Psychology & Aging*, *18*(3), 443–460.

Verhaeghen, P., Vandenbroucke, A., & Dierckx, V. (1998). Growing slower and less accurate: adult age differences in time-accuracy functions for recall and recognition from episodic memory. *Experimental Aging Research*, *24*(1), 3–19.

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition*, *13*(4), 304–319.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953–957.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*(8), 455–463.

Warren, L., & Marsh, G. (1979). Changes in event related potentials during processing of Stroop stimuli. *International Journal of Neuroscience*, *9*(4), 217–223.

Wascher, E., Reinhard, M., Wauschkuhn, B., & Verleger, R. (1999). Spatial S-R compatibility with centrally presented stimuli: An event-related asymmetry study on dimensional overlap. *Journal of Cognitive Neuroscience*, *11*, 214–229.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised (WAIS-R).* New York: Psychological Corporation.

Welford, A. (1965). Performance, biological mechanisms and age: A theoretical sketch. In A. Welford & J. Birren (Eds.), *Behavior, aging and the nervous system.* (pp. 3–20). Springfield, IL: Thomas.

West, R. (1999). Age differences in lapses of intention in the Stroop task. *Journals of Gerontology (B)*, *54B*(1), P34–P43.

West, R., & Baylis, G. C. (1998). Effects of increased response dominance and contextual disintegration on the Stroop interference effect in older adults. *Psychology & Aging*, *13*(2), 206–217.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272–292.

White, B. W. (1969). Interference in identifying attributes and attribute names. *Perception & Psychophysics*, *6*(3), 166–168.

White, I., & Wise, S. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research*, *126*(3), 315–335.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–101). New York: Academic Press.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors*, *30*(5), 599–616.

Willis, S. L. (1985). Towards an educational psychology of the older adult learner: Intellectual and cognitive bases. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 818–847). New York: Van Nostrand Reinhold.

Wingfield, A., Lindfield, K. C., & Kahana, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology & Aging*, *13*(2), 256–266.

Wise, S. P., & Murray, E. A. (2000). Arbitrary associations between antecedents and actions. *Trends in Neurosciences*, *23*(6), 271–276.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research*, *63*(3/4), 212–233.

Zacks, R., & Hasher, L. (1997). Cognitive gerontology and attentional inhibition: A reply to Burke and McDowd. *Journals of Gerontology (B)*, *52B*(6), P274–P283.

Appendix

I. Brinley regression parameters in 'multiplicative' tasks

Parameters of the Brinley plot regression are not only influenced by the true slowing factor(s), but also by the type of processing. This can be illustrated by generating points in Brinley space using models, where the true slowing factors as well as the true processing functions are set by the modeler. The following simulations assume identical slowing factors, however they turn out give dramatically different results, depending on the assumptions about the interdependency of different processing stages. Because the results are so different, they would suggest very different conclusions when interpreted using the 'standard' rationale that the Brinley slope approximates the slowing factor. A more coherent picture is obtained if separate regressions are estimated for each experiment, thereby eliminating between-experiment variance that is not of primary theoretical interest.

The modeling attempts were inspired by an attempt to meta-analytically integrate data from studies using Stroop-like tasks, in which 'early', e.g., stimulus-classification, and 'late', memory-related or episodic cognitive difficulty were orthogonally manipulated. The research question was whether old adults can be shown to have a specific deficit in 'episodic accumulators', which are thought to lead to an amplification of earlier difficulty effects. Here, it is argued that standard interpretations of graphical Brinley meta-analyses can be too much influenced by factors that vary between-experiment, but that are not the focus of the research.

Age-related slowing in peripheral sensuo-motor components is typically very small (but exists). Slowing in central information processing is larger. An estimation of the slowing proportion in a typical task is around 1.5, i.e. for every 100 ms difficulty effect in young adults' reaction times there is a corresponding 150 ms difficulty effect in old adults' reaction time. However, although the consistently linear relation in the Brinley plot indicates that general slowing of information-processing speed plays a major role in age-related slowing, the slowing factor seems to depend on aspects of the task at hand. For example, the slowing factor in 'coordinatively complex' tasks (that typically rely on working memory more than the average reaction time task) was found to be larger than 1.5, and similar results have been obtained for some facets of executive control. This constitutes evidence for the existence of different slowing factors for at least some cognitive processes.

Making a task more complex (difficult) will lead to some sort of latency increase. Typically this increase is caused by an increase in the duration of central, cognitive processes. Verhaeghen et al. (2002) describe additive and multiplicative complexity effects. If an extra processing stage is added to a task, then this is labeled additive complexity, because the manipulation will induce additive effects between the baseline and experimental conditions. If a complexity manipulation prolongs an existing stage in a task, then this is labeled multiplicative complexity, because it will induce multiplicative effects: time spent in the cognitive stage of the complex conditions will be a fixed ratio of time spent in the cognitive stage in the baseline condition. Verhaeghen et al. show that additive complexity leads to a pair of lines in state space, one for old and one for young, that are parallel to the diagonal, and either a single line or a pair of parallel lines in Brinley plots. Multiplicative complexity leads to slopes greater than unity in both of the scatter plots. Here I try to generalize their model for multiplicative complexity effects to a task where the duration of a 'late' cognitive stage is a function of the duration of an earlier cognitive stage.

In the following paragraphs, I try to determine some general properties of Brinley functions that follow from very simple models of processing. In particular, the models are linear, and only open-loop models are investigated, i.e. there is no feedback from later to earlier stages. Although both the open-loop and the linearity assumptions are not necessarily realistic, they are used as a first approximation. The goal is to answer questions about Brinley functions such as the following: What is the dependency of the slope of a Brinley function on the slowing factors of component processes? How does the slope of a Brinley plot depend on the connection between cognitive stages? For example, the stages could be serial and independent in task A, or they could be serial, but multiplicatively connected in task B, such that the duration of a later stage C_2 is positively correlated with the time spent in an earlier stage C_1.

General slowing

To introduce the general line of argumentation, I will start with a very simple model, where reaction time is a sum of a peripheral sensorimotor processing stage and central cognitive processing stage. The model will then be expanded by the insertion of a second and a third cognitive stage.

Take a very simple model of task performance, $RT = S + C$, where reaction time (RT) is determined by the durations of two additive stages, one sensorimotor (S), and one cognitive (C). Experimental difficulty manipulations are assumed to affect only the cognitive stage. Furthermore, based on empirical data, we assume that sensorimotor slowing is negligible, and that proportional age-related slowing affects cognitive processes. That is, we assume a slowing factor $\lambda \geq 1$ for C, and a lack of slowing for S. Formally, the model assumptions can be written as

$$\begin{aligned} RT_{young} &= S + C \\ RT_{old} &= S + \lambda C \quad . \end{aligned} \qquad (20)$$

By assumption, only C is variable because C, not S, is affected by difficulty manipulations. To observe how young and old adults' reaction times change with C, we determine the derivatives

$$\frac{dRT_{young}}{dC} = 1 \quad , \quad \frac{dRT_{old}}{dC} = \lambda \quad .$$

We are interested in the relation of the parameters of our simple model and parameters of the Brinley function

$$RT_{old} = \beta_1 RT_{young} + \beta_0 \tag{21}$$

with a derivative of

$$\frac{dRT_{old}}{dRT_{young}} = \beta_1 \quad . \tag{22}$$

If we equate $\frac{dRT_{old}}{dRT_{young}} = \frac{dRT_{old}}{dc}\frac{dc}{dRT_{young}}$, we obtain

$$\beta_1 = \lambda \quad . \tag{23}$$

Inserting this result and the model equations (20) into (21) yields

$$\begin{aligned} S + \lambda C &= \lambda(S + C) + \beta_0 \\ \beta_0 &= S(1 - \lambda) \quad . \end{aligned} \tag{24}$$

Thus in our simple model, a cognitive difficulty manipulation that affects only C will lead to a Brinley plot of the form

$$RT_{old} = \lambda RT_{young} + (1 - \lambda) S \quad . \tag{25}$$

We note that in (25), the intercept b_0 of the Brinley function depends on the slowing factor, which can explain the empirical fact of a negative correlation of b_0 and b_1 between Brinley plots of different tasks (see also Ratcliff et al., 2000). Since $\lambda > 1$, this dependency can also explain why negative intercepts are typically found in Brinley plots. An example using a 'typical' duration of 200 ms for sensorimotor processing and a 'typical' slowing factor of $\lambda = 1.5$ for the central component produces a Brinley function with a slope of 1.5 and an intercept of -100 ms, which is in good agreement with empirical observations.

Hence this very simple model can already explain some of the Brinley plot regularities. It cannot, of course, explain the empirical fact of Brinley functions with different slopes. In an attempt to account for these, let us expand the model to include additional cognitive stages.

Process-specific slowing

In the following, I will consider Brinley plot predictions of a slightly extended, yet still simple and general processing model. Again it is assumed that reaction time is the sum of the time spent in different stages: Response time for young adults, RT_y, is the sum of the duration of a sensory stage, S, and three cognitive stages, $C_i, i = 1\ldots3$. For convenience, I will label the three cognitive stages 'early processes', 'memory', and 'response preparation'. To generate response time for old adults, RT_o, each of the stages is associated with its own age-related slowing factor λ_i.

Here is the general model:

$$\begin{aligned} RT_y &= S + C_1 + C_2 + C_3 \\ RT_o &= S + \lambda_1 C_1 + \lambda_2 C_2 + \lambda_3 C_3 \end{aligned} \tag{26}$$

In many situations, early and late cognitive stages are not independent. In particular, this will be the case in conflict tasks, e.g., the Stroop task. For conditions with high episodic demands, it is assumed that effects of 'early' manipulations, e.g., color-word (in-)congruency, are amplified by episodic accumulators, located at the cognitive stage. Episodic accumulators are needed when the task requires arbitrary rules, e.g., in the case of arbitrary stimulus-response mappings. To model this, let the duration of episodic processes be a multiple of the duration of earlier processes:

$$C_2 = b_1 C_1 \tag{27}$$

This can also be used to model conditions with low episodic demands, e.g., when a compatible stimulus-response mapping is used. Here, episodic accumulators can be bypassed, i.e., $b_1 = 0$.

Arbitrary conditions

If we insert (27) into the general model (26), we obtain

$$\begin{aligned} RT_y &= S+(1+b_1)C_1+C_3 \\ RT_o &= S+(\lambda_1+b_1\lambda_2)C_1+\lambda_3C_3 \end{aligned} \tag{28}$$

This leads to a Brinley-Plot with the slope

$$\begin{aligned} \beta_1 &= \frac{\partial RT_o}{\partial RT_y} \\ &= \frac{dRT_o}{dC_1,dC_3}\frac{dC_1,dC_3}{dRT_y} \\ &= \frac{(\lambda_1+b_1\lambda_2)dC_1+\lambda_3dC_3}{(1+b_1)dC_1+dC_3} \end{aligned} \tag{29}$$

The intercept of the Brinley regression is given by

$$\begin{aligned} \beta_0 &= RT_o-\beta_1RT_y \\ &= S+(\lambda_1+b_1\lambda_2)C_1+\lambda_3C_3-\beta_1(S+(1+b_1)C_1+C_3) \\ &= (1-\beta_1)S+(\lambda_1+b_1\lambda_2-\beta_1(1+b_1))C_1+(\lambda_3-\beta_1)C_3 \end{aligned} \tag{30}$$

From equation (29), we see that if $dC_3 \neq 0$ between experiments, then the Brinley slope will not exclusively depend on the parameters under investigation, λ_1 and λ_2, but also be influenced by the theoretically less interesting between-experiment changes of C_3.

On the other hand, if $dC_3 = 0$, i.e., no additional process enters with a change in conditions (likely within an experiment, but also possible between experiments), then

$$\beta_1 = (\lambda_1+b_1\lambda_2)/(1+b_1) \tag{31}$$

That is, if there is no change in the duration of response preparation processes, then the slope is a weighted mean (weighted by b_1) of early and late/episodic slowing factors.

Compatible conditions

Let us now consider the situation in which episodic accumulators can be bypassed. This can be formally expressed by letting $b_1 = 0$. In this case, model (28) reduces to the following simple model:

$$\begin{aligned} RT_y &= S+C_1+C_3 \\ RT_o &= S+\lambda_1C_1+\lambda_3C_3 \end{aligned} \tag{32}$$

with the Brinley plot slope

$$\begin{aligned} \beta_1 &= \frac{\partial RT_o}{\partial RT_y} \\ &= \frac{\lambda_1dC_1+\lambda_3dC_3}{dC_1+dC_3} \end{aligned} \tag{33}$$

Here, the Brinley regression slope is the mean of the slowing factors, weighted by the effect sizes.

The intercept is given by equation (30), which in the case of $b_1 = 0$ can be simplified, obtaining

$$\beta_0 = S(1-\beta_1)+C_1(\lambda_1-\beta_1)+C_3(\lambda_3-\beta_1) \tag{34}$$

When episodic demands are absent, both slope and intercept depend on early and response-related slowing factors, however, the cognitive slowing factor is (of course) eliminated.

Equivalence of 'peripheral' cognitive slowing

If additionally $\lambda_3 = \lambda_1$ (as is often assumed), i.e., early cognitive processes (e.g., stimulus classification) and response preparation are associated with the same slowing factor. In this case, the Brinley plot predictions for conditions in which episodic accumulators are bypassed can be further simplified. First, for the slope, by replacing λ_3 with λ_1 in equation (33) we obtain

$$\beta_1 = \frac{\lambda_1 dC_1 + \lambda_1 dC_3}{dC_1 + dC_3} = \lambda_1 \tag{35}$$

Second, by replacing λ_3 with λ_1 and β_1 with λ_1 in equation (34), for the intercept we obtain

$$\beta_0 = S(1 - \lambda_1) \tag{36}$$

Thus, if episodic accumulators can be bypassed and the slowing factors associated with peripheral cognitive processing are the same, then the simple Brinley model (25) can be reproduced.

If however, episodic accumulators cannot be bypassed, then even if $\lambda_3 = \lambda_1$, the equations are fairly complex. In particular, they are still affected by between-experiment changes in the sizes of 'peripheral' effects, dC_1 and dC_3. For example, equation (29) becomes

$$\begin{aligned}\beta_1 &= \frac{(\lambda_1 + \lambda_2 b_1)dC_1 + \lambda_1 dC_3}{(1 + b_1)dC_1 + dC_3} && (37)\\ &= \frac{\lambda_1(dC_1 + dC_3) + \lambda_2 b_1 dC_1}{(dC_1 + dC_3) + b_1 dC_1} && (38)\end{aligned}$$

As an example, I will now consider the application of the general model for experiments that follow the rationale of the experiments reported in this dissertation. Two instances of the model will be compared, in which the degree of involvement of a third stage varies between experiments (or tasks). Both are used to investigate a situation in which a Brinley analysis combines data from experiments that use an orthogonal manipulation of early difficulty (e.g., Stroop condition) and 'episodic difficulty' (or short-term memory demands). The examples only differ in the assumptions about the third-stage slowing factor.

Both examples assume that within a given task (or experiment), episodic accumulators (on the late, cognitive stage C_2) will lead to a proportional amplification of earlier difficulty effects (originating at stage C_1). The two examples differ in the assumptions about the slowing of processes that vary between experiments. In the first example, E1, it is assumed that the age-related slowing factor for these processes, λ_3, is equal to the 'episodic' slowing factor, λ_2. In the second example, E2, it is assumed that age-related slowing of processes varying between experiments is small, with the same slowing factor as early cognitive processes $\lambda_3 = \lambda_1$. (Both examples assume that processes varying between tasks do not interact with the 'early' difficulty manipulation.) Within each example, experiments differ with respect to the duration of stage C_3. Three different experiments were simulated, with C_3 durations of 0, 100, and 250 ms. For all experiments, 'compatible' ($b_1 = 0$) and 'arbitrary' ($b_1 = 0.75$) conditions were simulated, using the same levels of early difficulty effects (C_1).

The result, Brinley plots according to examples E1 and E2, is shown in figures 21 and 22. The 'true' slowing factors and other parameters used to generate these plots are

$\lambda_1 = 1.05$ slowing of early cognitive processing

$\lambda_2 = 3.0$ slowing of late cognitive processing (episodic)

λ_3 slowing of response preparation processes

$$\lambda_3 = \begin{cases} 3.00 & \text{E1: } \lambda_3 = \lambda_2\text{, same as episodic slowing} \\ 1.05 & \text{E2: } \lambda_3 = \lambda_1\text{, same as early cognitive slowing} \end{cases}$$

b_1 factor characterizing the dependency of late on early processing duration

$$b_1 = \begin{cases} 0.75 & \text{arbitrary: 'episodic accumulators' are needed} \\ 0.00 & \text{compatible, 'episodic accumulators' can be bypassed} \end{cases}$$

$S = 200$ duration of peripheral processes, assumed to be unaffected by age-related slowing

$C_1 = \{0, 50, 120\}$ duration of early cognitive processes

$C_3 = \{0, 100, 250\}$ duration of response preparation processes

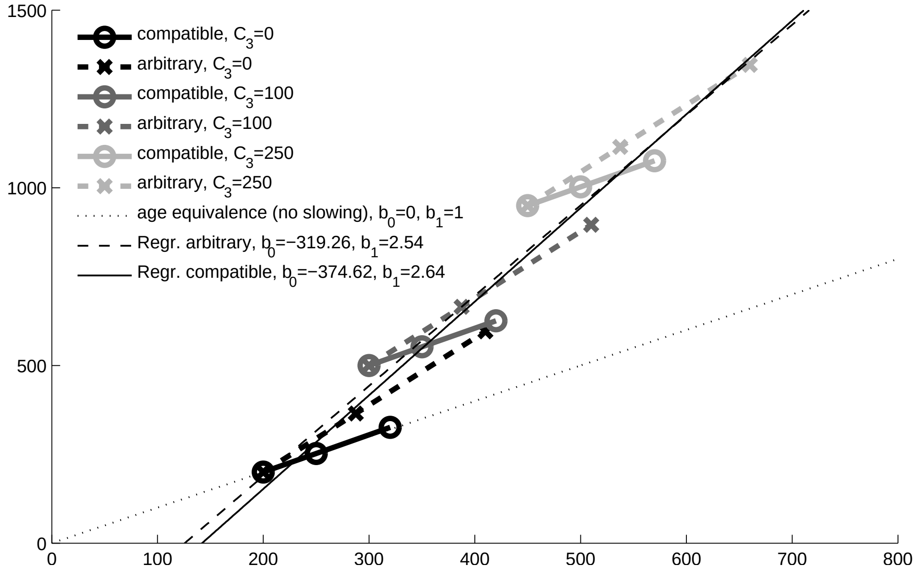


Figure 21. Example E1: $\lambda_3 = \lambda_2$. Age-related slowing for processes varying between experiments is equal to slowing of episodic memory processes, $\lambda_2 = 3.0$.

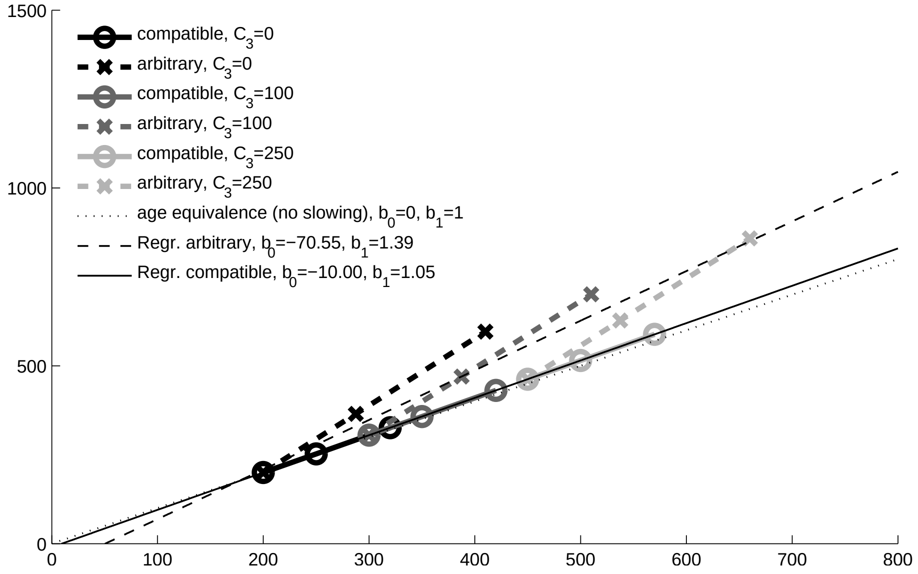


Figure 22. Example E2: $\lambda_3 = \lambda_1$. Age-related slowing for processes varying between experiments is equal to slowing of early cognitive processes, $\lambda_3 = \lambda_1$.

Apparently, the (global) Brinley slopes are rather different, depending on the differences between experiments. However, let us assume that the experiments were selected for the meta-analysis because the analyst was interested in the commonalities, not in the differences between experiments. Commonalities, in the current examples, stem from the fact that the same processes C_1 and C_2 were investigated in all experiments. The overall Brinley slopes largely depend on C_3 and the associated the slowing factor λ_3. Therefore, conclusions drawn from the overall slope appear to be flawed.

On the other hand, if separate regressions are fit for each experiment, we see that the slope is consistent across experiments, regardless of the value of λ_3. Within each experiment, the slopes are $\beta_1 = 1.89$ for arbitrary mappings (thick, dashed lines), and $\beta_1 = 1.05$ for compatible mappings (thick, solid lines). These same factors are obtained in both examples E1 and E2. Thus, if meta-analytical integration of age-related slowing data is done to investigate the slowing of some 'multiplicative' central process like episodic memory, then the standard Brinley plot approach should be replaced by an approach that fits separate regression lines for each experiment.