Aspects of Declarative Memory Functioning in Adulthood: Cross-Sectional and Longitudinal Studies

Michael Rönnlund

Department of Psychology, Umeå University, Sweden
2003
Aspects of Declarative Memory Functioning in Adulthood: Cross-Sectional and Longitudinal Studies

Michael Rönnlund
ABSTRACT


The general objective of the thesis was to examine aspects of declarative memory functioning across the adult life span. The four papers were based on data collected as part of the Betula Prospective Cohort Study (Nilsson et al., 1997) and included large-scale population-based samples of participants in the age range 35 to 90. In study I and study II the possibility that age differences in episodic memory may be compensated for by provision of encoding support in the form of enactment was investigated, using free and cued recall and recognition portioned into components of recollective experience as the dependent measures. In Study III, unitary, two-, and multi-factorial models of declarative memory were compared and age-invariance was tested for. In Study IV cross-sectional age differences were contrasted with five-year longitudinal changes on aggregate measures of episodic and semantic memory within age groups ranging from 35 to 85 years. The results of Study I and Study II demonstrated that enactment constitutes an effective form of encoding support, but that the age differences generalize across this form of encoding support. Study II indicated that most of the age-related variance in recognition and levels of recollective experience following enacted and non-enacted encoding was shared by a measure of processing speed. Study III confirmed that a two-factor model of declarative memory (episodic and semantic memory) yields superior fit as compared with a unitary model of declarative memory. However, the best fitting model was a six-factor model with recall and recognition (episodic memory) and knowledge and fluency (semantic memory) as first-order factors. Episodic and semantic memory were found to be associated with different patterns of age-related differences, with some indications that the first-order factors show differential age-related patterns, indicative of variability within the episodic and semantic memory domains. The results of Study IV showed that cross-sectional data may give a false impression of decline for adults in the age range 35-60 years for episodic memory. There was no evidence of time-related decline within these age groups, even though practice effects were taken into account. However, past this age, substantial time-related decline was observed for the older adults, in line with cross-sectional data. Semantic memory performance tended to improve across time for the younger groups, but decline in old age, although the magnitude of this decline was less pronounced than for episodic memory. Cohort differences in education may be one important factor underlying the discrepancy between the cross-sectional and longitudinal aging patterns, both in the case of episodic and semantic memory. In conclusion, the result of the present studies show that age-related functional losses occur in forms of declarative memory, especially episodic memory, but that the onset of decline does not begin until old age.

Key words: aging, episodic memory, semantic memory, encoding support, cross-sectional, longitudinal.
ACKNOWLEDGEMENT

The completion of my thesis was made possible by the support of numerous persons. I would especially like to thank the following persons.

First and foremost, my supervisor during most of my time as a doctoral student, Lars Nyberg, for providing me tremendous support and encouragement, always willing to assist me in various matters, from reading my first lousy drafts to helping me formulate the conclusions.

I would also like to thank Lars-Göran Nilsson, my supervisor during the initial stage of my graduate studies and co-author of all the empirical studies, for introducing me to the psychology of memory and the Betula study and for serving as source of inspiration for me and the other members of the research project ever since. Special thanks also to Lars Bäckman, co-author of two studies.

I also wish to thank my colleagues in the Betula study with which I spent numerous working days during the data collection periods and without whom several dear episodic memories would not have been acquired. The list of past and current members of the Betula staff includes too many persons to name at present, but the “returnees”, including Gunilla Smedberg-Åman, Ann-Louise Söderlund, and Maud Widing deserve special attention. This holds also for Martin Lövdén who I had the opportunity to get to know at “Time 3” and with whom I have cooperated with at the level of scientific enquiry ever since, with one of the included papers, and several papers, as a result. I am happy to have him as a friend (and to beat him in chess on occasions).

In addition, I would like to thank fellow doctoral students, some of which achieved a higher level of processing speed than me and completed their theses a while ago, including Arne Börjesson, Reza Kormi-Nouri, Michael Gruber, and Betula fellows Hedvig Söderlund, Jonas Persson, and those who are still in progress, including Anna Sundström and Petter Marklund. As regards the staff at the department of psychology, special thanks to those who taught courses at the graduate level in particular, including Anders Böök, Ingvar Lundberg, Timo Mäntylä, and Bo Molander, the latter who provided constructive critique on an earlier draft of this thesis together with Agnetha Herlitz. Thanks to you.

Finally, I would like to thank all of the participants in the Betula study, a few of which I had the opportunity to meet personally during the data collection waves.

Umeå, October, 2003
Michael Rönnlund
LIST OF PAPERS

The thesis is based on the following studies, referred to in the text by their roman numerals.


INTRODUCTION AND BACKGROUND

Efficient encoding, storage, and retrieval of information is essential to cope with most things in life, including the ability to walk, retain ones identity, communicate with other people, and mentally re-live past experiences. Most elderly people report that their ability to remember has gotten worse since they were younger (Ryan, 1992). The results from a wealth of studies have confirmed this assertion (Kausler, 1994). Even though concepts such as primary (Busse, 1969) or “normal” (Schock et al., 1984) aging are fuzzy (see Birren & Shroots, 1996 for an overview of different aging concepts), this pattern of age-related deficits is observed in the relative absence of disease. At the same time, research conducted during the last few decades converges on the notion that memory, rather than being a unitary construct, is decomposable into different systems or components of processing. In a related vein, the literature provides a great deal of support of the view that not all aspects of memory “age alike”; some may be subject to considerable functional losses, whereas others are relatively spared by the aging processes. Apart from determining the relative age-sensitivity among various memory measures, an important quest in research on aging and memory has been to appreciate the extent to which deficits in age sensitive forms of memory may be remedied following provision of cognitive support. The empirical studies of the present thesis addressed both of these broad issues and they will be dealt with in more detail in subsequent sections.

Even though numerous studies devoted to the effects of aging on memory have been published during the last decades, and significant gains in knowledge concerning the boundary conditions of adult age-related effects have been made, the majority of the studies suffers from two shortcomings: (a) they involved extreme age-group comparisons, that is contrasted a group of young adults often in their 20s with a group of old adults, typically in their 60s or 70s, and (b) they were cross-sectional, that is, they inferred ontogenetic change from age differences observed at a single point in time. As a consequence, the question of whether life-span changes in various memory functions progress in a continuous or discontinuous fashion can not be fully appreciated from the bulk of data. In the cases were a wider age range was included the effects due to the specific conditions under which people grew up (cohort factors) are difficult to distinguish from intra-individual changes.

The empirical studies of the thesis were based on cross-sectional (Study I, II, and II) and longitudinal data (Study IV) from population-based samples of adults ranging in age from 35 to 90 years, data emanating from the Betula prospective cohort study (Nilsson et al., 1997). To provide a background of the issues addressed by the individual studies, major classificatory schemes of memory and the ways that different forms of memory have been measured will first be introduced, with focus on what will be referred to as declarative memory, and a sub-form known as episodic memory in particular, as this form of memory was specifically studied in all four of the empirical studies. A select review of the literature on aging and memory pertaining to the central themes of the thesis will thereafter be given. Finally, the reader is provided with an overview of the design of the
Declarative and non-declarative memory

Over the last two or three decades much effort has been devoted to the issue of how the variety of memory tasks assumed to reflect long-term memory should be classified. One well-established distinction is that between declarative and non-declarative memory. This distinction draws on the separation of two forms of knowledge (knowing that versus knowing how) made by Ryle (1949). Specifically, declarative memory refers to encoding, storage, and retrieval of propositional knowledge that is under conscious control (knowing that), including memory of personal events and factual knowledge (Cohen & Squire, 1980; Squire, Knowlton, & Musen, 1993). The term non-declarative memory (knowing how), on the other hand, is used to denote memory of various nonverbal behaviors or skills, not readily expressible in verbal form, in case of which retention is not necessarily accompanied by conscious recollection of the original learning situation (cf. implicit memory, Schacter, 1987). A variety of forms of learning, including the gradual acquisition of motor and cognitive skills, perceptual priming, classical conditioning, and non-associative learning are subsumed by the term non-declarative memory (Squire & Zola-Morgan, 1991).

Evidence that performances on declarative and non-declarative measures are dissociable was first provided from studies of amnesic patients suffering from damage to the medial temporal lobes. These patients typically show severe deficits as compared with healthy controls in declarative memory tasks, but exhibit normal, or close to normal, level of performance on various measures of non-declarative memory (e.g., Cohen & Squire, 1980; Brooks & Baddeley, 1976; see Squire, 1994 for a review). Such findings led to the proposal that declarative memory is critically dependent on the medial temporal lobe system (including the hippocampus), whereas non-declarative memory functioning is dependent on various non-temporal brain regions (Prull, Gabrieli, & Bunge, 2000).

In a different taxonomy of long-term memory proposed by Tulving and Schacter (1994), non-declarative memory is fractionated into two subsystems. One of these subsystems (the perceptual representation system) is proposed to handle object identification. This system is assumed to account for so called repetition-priming phenomena, where previous presentation of a stimulus (e.g. a word) facilitates subsequent identification of the stimulus presented in a degraded form (e.g. a word-fragment). This facilitation may be observed quite independent of recollection of the original event or stimulus. Another memory system, procedural memory, is assumed to subserve learning and retention of various kinds of motor and cognitive skills (see Nyberg & Tulving 1996).

It is beyond the scope of the present thesis to discuss aging effects in non-declarative forms of memory. Suffice is to note that meta-analytic evidence and literature reviews tend to reveal negative age-related effects in several tasks assumed to tap these forms of memory (LaVoie & Light, 1994; Kausler, 1994). However, the magnitude of the age-related deficits is generally much smaller than in measures assumed to tap aspects of declarative memory and often not detectable in single studies. Thus, capacities of learning
and retention that can be subsumed by the term non-declarative (or implicit) memory, appears to be relatively age resilient.

**Episodic and semantic memory**

Whereas the proposal of a declarative memory system implies that memory of facts and events, at least in part, are subserved by common mechanisms, Tulving (1972, 1983) proposed a distinction between recollection of personally experienced events and memory of general knowledge, which he termed episodic and semantic memory, respectively. Tulving noted that retrieval of episodic and semantic memories is associated with different kinds of awareness. Specifically, retrieval of episodic memory may be characterized by autonoetic, or self-knowing, consciousness, whereas semantic memory is characterized by noetic consciousness which lack reference to the personal past. Thus, episodic memory is assumed to be responsible for mental revival of happenings in your personal past, whereas semantic memory is concerned with common knowledge, devoid of personal reference. To provide an example, successful retrieval of last Christmas Eve likely brings to mind details such as what you did and where you were, accompanied by a feeling that this event belongs to your personal past. By contrast, answers to the questions: “What quadruped animal wags its tail to friends and barks to strangers?” or, “What is the Capital of Egypt?” are likely retrieved separate from recollection of the personal circumstances at time of the original encoding. Tulving regarded episodic and semantic memory as different systems but has pointed out that interactions occur. Specifically, semantic memory has been viewed as the more basic system and the direction of influence between episodic and semantic memory has been hypothesized to be such that semantic memory primarily influences episodic memory (Tulving, 1985a).

In laboratory settings episodic memory is typically simulated by presenting a series of stimuli or items (e.g. words or pictures) to be remembered for purpose of a later test. In some cases, the participants are requested to reproduce, or recall, the studied items (e.g., by writing them down in the case of words), either with or without some cue (e.g. an associate of the word presented at study) being presented at time of retrieval. In other cases, the participants are requested to judge whether each of the items in a test list are old or new. Some of these items were studied previously (targets) and some were not (distractors). This is an example of a recognition test. The foregoing episodic measures concern retrospective memory, that is, situations in which the memory search is directed to past events. Memory for tasks intended to be performed in the future, or prospective memory (e.g., Maylor, Smith, Della-Sala, & Logie, 2002), may also be subsumed under the heading of episodic measures. As the present research focused on retrospective memory, the literature on prospective memory will not be discussed further. A variety of markers of semantic memory have been used, including tests of general knowledge or vocabulary, lexical decision tasks, and tasks requiring rapid reproduction of verbal materials (e.g., word fluency tests).
Task dissociations

A (single) dissociation is established if a variable has a different effect on measures assumed to reflect separate constructs. Several kinds of dissociations have been taken to indicate that episodic and semantic memory constitute functionally and neurologically separate systems (Nyberg & Tulving, 1996). A functional dissociation is observed if an experimental variable has a different effect on measures assumed to reflect two different constructs. To take an example pertaining to the episodic-semantic distinction, one type of encoding manipulation may have a positive effect on episodic recall, but no effect on the ability to produce instances from the categories into which the studied materials are divided (category association), a task assumed to reflect semantic memory (Nyberg & Nilsson, 1995). A pharmacological dissociation is observed if a drug-induced state has a different effect on measures of two different constructs. To this point, Roy-Byrne et al. (1987) found that diazepam had a negative effect upon two episodic measures (recognition and recall) but no effect on a measure of semantic memory. A finding of a differential effect of brain damage on two tasks provides another potential source of evidence that these tasks reflect partially different underlying structures or mechanisms. Results showing that patients suffering from Korsakoff’s syndrome are impaired on measures of recall or recognition, despite normal performance on various semantic measures (Weingartner, 1983; Shimamura & Squire, 1989) have thus been taken as support of the view that episodic and semantic memory constitute neurologically separable forms of memory. Yet another type of dissociation concerns developmental effects. This is of primary interest in the present thesis and the aging patterns observed on measures of episodic and semantic memory will be discussed later.

Patterns of brain activation observed in healthy participants performing episodic and semantic tasks should provide important insights as to the differences and similarities between episodic and semantic memory. With regard to activity in prefrontal regions of the brain, retrieval of episodic and semantic information has been found to show distinct patterns. Specifically, retrieval of episodic information is associated with activation of right prefrontal cortex, whereas retrieval of semantic information is associated with activation in the left prefrontal cortex (e.g., Nyberg et al., 1996). In addition, there are indications that the medial temporal lobe, in particular hippocampus, is related to episodic memory performance (Cabeza & Nyberg, 2000) and that the left lateral temporal lobe is involved in semantic memory performance (Dalla Barba et al., 1998). Left inferior prefrontal cortex seems to play a role in both forms of memory (Dalla Barba et al., 1998; Cabeza & Nyberg, 2000). Thus, neuroimaging data reveal differences between measures of episodic and semantic memory but also some similarities.

It should be noted that not all researchers have agreed on the usefulness of the distinction between episodic and semantic memory, in particular the view that various types of dissociations are indicative of different memory systems (e.g., McKoon, Ratcliff, & Dell, 1986). In general, some researchers have instead adopted the view that dissociations between measures taken to reflect different memory systems may be attributed to differential involvement of various processes that operate at encoding and
retrieval in different tests (e.g. Jacoby & Dallas, 1981, Blaxton, 1995). As noted by others, the processing and systems views are probably best regarded as complementary. Specifically, memory processes may be regarded to operate within structures (systems) and systems may be regarded as components of processing. Researchers have noted that it is difficult to distinguish between system- and components of processing oriented explanations at the current time (Parkin, 1999).

Of more fundamental concern, most studies contrasted retrieval of recently encoded episodic events and tests requiring retrieval of old knowledge. This, it may be argued, leaves open for the possibility that episodic and semantic memory are subserved by common encoding processes despite the fact that retrieval of (new) episodic and (old) semantic memories are dissociable. It may also be argued that propositional information gradually become devoid of spatial-temporal context with repeated exposure, such that semantic memories represent the residue of many episodes (e.g., Baddeley, 2002). In the case of amnesia, studies have indeed observed that patients may be able to retrieve old semantic (and episodic) memories, together with problems of acquiring new semantic information (e.g., Baddeley & Wilson, 1986). On the other hand, cases of anterograde amnesia in which acquisition of new semantic information, including post-morbid vocabulary and facts, may take place despite severe inability to remember new episodic information have been reported (Kitchener, Hodges & McCarthy, 1998).

In sum, several kinds of dissociations have been taken to support that episodic and semantic memory constitute separate memory systems, including evidence that partly different brain areas are involved when participants retrieve episodic and semantic memories. However, the issue of whether episodic and semantic memory are best conceived of as separate systems or as differing along a continuum from recently acquired to old memories does not appear to be fully resolved.

Episodic and semantic memory: separate ability factors?

From the view that episodic and semantic memory constitute different, although inter-related, forms of declarative memory, it may be expected that a pattern of strong relations among measures of the same construct (e.g., episodic or semantic memory) is observed and that the two classes of measures are less than perfectly related. This is so, regardless of whether the evidence is interpreted in light of a memory systems view or a components of processing view. In other words, a separable ability structure should be discernible from performances on a battery of tests that include multiple markers of the episodic and semantic memory. Surprisingly little research has approached the issue of whether proposed memory constructs such as episodic or semantic, perceptual, and procedural memory and so forth are separable from an individual differences, or psychometric, approach. This likely reflects the fact that research devoted to the issue of multiple memory forms was rooted in the experimental and neuropsychological traditions.

As early as in the late 1930s one can find attempts to distinguish intellectual factors that bear relevance to the issue of whether tests involving retrieval of recently acquired information (episodic memory) and tests of verbal knowledge (semantic
memory) constitute separable abilities. A study by Thurstone (1943/1938) used exploratory factor-analytic methods to sort out ability factors from a broad set of tests. The initial results were taken to support the existence of seven major primary mental abilities. Of primary interest here, an ability factor labeled “associative memory”, mainly reflecting tests that we would call episodic measures was identified. Furthermore, two factors reflecting performances on tests typically considered as measures of semantic memory, verbal meaning, on the one hand, and word fluency on the other, were distinguishable from point of these analyses (we shall return to the possibility that semantic as well as episodic memory can be fractionated further into sub-factors).

A study by Underwood, Baruch, and Malmi (1978) appears to be the first study that systematically investigated interrelations between memory measures labelled as episodic and semantic. The results suggest that performances on the two classes of tests may be quite unrelated. The low correlations observed between measures of episodic measures on the one hand and semantic measures (e.g., a test of vocabulary and spelling) on the other, is however, somewhat difficult to interpret in light of relative lack of association among the semantic measures (as was also noted by McKoon et al., 1986).

A few subsequent studies have approach the issue of the factorial structure of episodic and semantic memory measures. Mitchell (1989), for example, found support for a two-factor solution consistent with the episodic-semantic distinction using a variety of markers of the two test types. Whereas the foregoing studies were conducted using exploratory factor analysis, a more recent study by Nyberg (1994) instead used confirmatory factor analyses to test the viability of the episodic-semantic distinction, based on data for adults in the age range 35 to 50 years. The latter type of approach provides the advantage that it allows for a comparison of the fit of competing theoretical models. The results suggested that a two-factor model, fractionating declarative memory into semantic and episodic factors, provided a better account of the covariance relation among tests than a unitary, or single-factor, model.

In sum, as judged from the results of the few extant studies, episodic and semantic memory measures load on different ability factors.

**Cross-sectional aging patterns in episodic and semantic memory measures**

Several lines of evidence suggest that measures of episodic memory represent reflect partly different memory abilities. We now turn to aging patterns discernable in measures of these two constructs. For the moment, cross-sectional age comparison will be considered. The possibility that these patterns differ from those emergent from longitudinal studies will be addressed later.

**Episodic memory**

Episodic memory is typically regarded as highly age sensitive. This is based on a wealth of studies showing that groups of older adults are at disadvantage as compared to younger
adults in memory of a variety materials, including verbal (e.g., words, digits, sentences, and text) and nonverbal (e.g. faces and objects) materials. The result of negative age differences, it should be noted, appears to generalize across presumably more ecologically relevant materials than those typically used in laboratory studies, such as medication labels (Allaire & Marsiske, 1999), news (Frieske & Park, 1999), and mini-golf shots (Molander & Bäckman, 1990), to give some examples.

In addition, age-related deficits in episodic tests have been observed across various procedural variations. To provide a few examples, older adults perform worse than younger adults following intentional as well as incidental encoding, and under self-paced as well as time-restricted learning conditions (see Verhaeghen, Marcoen, & Goosens, 1993 for a meta-analysis).

Furthermore, the age-related pattern appears to generalize across several major demographic variables. Salthouse (1991) summarized a number of cognitive aging studies, including studies of memory that concerned within-occupation contrasts. He found little evidence for differential aging-patterns depending on occupational status (i.e., the different within-occupation contrasts revealed similar magnitudes of age-differences). To mention another major demographic variable, gender, data seem to reveal equal magnitudes of age-related deficits in episodic tasks for men and women (for a meta-analysis, see Meinz & Salthouse, 1998).

Do the depicted age-related memory losses progress in a continuous or discontinuous fashion from early to late adulthood? In the past decade a few large-scale studies covering a wider age-range have been published. In general these appear to be suggestive of a linear age-performance relationship. To take an example, cross-sectional findings by Park et al. (2002) indicate that the onset of decline may begin as early as in the 20s or 30s on episodic measures followed by a gradual decline throughout the remainder of the life span. As judged from these data, the magnitude of decline is considerable. At age 60 the average individual will be expected to perform as much as one standard-deviation unit below peak level (a figure that agrees reasonably well with estimates from meta-analytic studies, e.g., LaVoie & Light, 1994; Verhaeghen & Salthouse, 1997), and, at age 80, two standard deviation units below peak level.

The foregoing patterns concern measures of recall. Age differences tend to be smaller on measures of recognition. However, in contrast with some early studies, indicating that measures of recognition are associated with negligible age effects (e.g. Schonfield & Robertson, 1966), a large number of studies reveal age differences that are at least moderate in magnitude also in recognition tests (> .5 standard-deviation units, cf. Spencer & Raz, 1995).

Interestingly, the finding of a (partial) dissociation between recall (highly age sensitive) and recognition measures (less age sensitive) mirrors the pattern observed in amnesic patients (e.g., Hirst, Jonsson, Kim, Phelps, & Volpe, 1986), showing recall deficits in the relative absence of recognition deficits. Also, word frequency has a differential effect on recall and recognition (Gregg, 1976) and intentionality to remember often has a larger effect on recall than on recognition (e.g., Koriat & Feuerstein, 1976), providing examples of functional dissociations. Finally, neuroimaging data have indicated
partly different activation patterns in recall and recognition performance (Cabeza et al., 1997). Based on such observations, a possibility that was investigated further in one of the empirical studies is that recall and recognition reflect distinguishable ability factors and that there these factors vary with regard to the magnitude of adult age differences.

**Semantic memory**

The definition of semantic memory as the individuals’ store of world knowledge is very similar to what has been referred to as “crystallized intelligence”, that is knowledge acquired through acculturation (Horn & Cattell, 1966, 1967). Measures of vocabulary, fact recall, and comprehension, which reflect semantic knowledge, have a long history in mental testing (often labelled as “verbal ability measures”). The studies in which they were included were often normative in orientation. As consequence, there is relatively much cross-sectional data spanning a wider age range available for these measures.

In general, the patterns of age differences observed on these tests are quite different from that observed on measures of episodic memory. Across the young (20 years) to young-old period (60 years), cross-sectional data indicate stability of, or increase (especially when educational differences are controlled for) in performance in test of vocabulary and verbal meaning (e.g., Wechsler, 1981; Birren & Morrison, 1961). In the period from young-old age to late senescence some studies report negligible differences (e.g., Christensen, 2001) or even a tendency suggestive of an age-related increment in performance (Park et al., 2002). At the same time, other studies are indicative of a deficit in old age also on such measures (e.g., Berkowitz, 1953; Lindenberger & Baltes, 1994). However, in these cases, it should be noted that the magnitude of depicted decline past peak-level age is typically smaller than for episodic measures. Moreover, much of the deficits in old age may be out-weighed by the aforementioned tendency of increases up to middle age (e.g., Kaufman & Horn, 1996).

Results from studies using other types of measures also suggest that old age is associated with relatively preserved semantic memory. Semantic memory is often conceived of as a network where concepts are stored as nodes and where spreading activation within the network is a basic mechanism underlying retrieval of information from memory (e.g., Collins & Loftus, 1975). To investigate the extent to which aspect of storage and retrieval from this network differ in younger and older adults, latency-measures and word association norms are informative. In general these results provide much evidence that that storage and retrieval of semantic information is little affected by the aging process. That is, several studies reveal age invariance with regard to associations among words and spared semantic priming (facilitation of retrieval of a concept by a related one). At the same time, the spread of activation, and hence the time to retrieve words is generally slower in older adults. Older adults tend to have problems of lexical access, as judged from increased naming errors and are more prone to experience TOT (tip-of-the-tongue) states (e.g., Maylor, 1990; for a review, see Light, 1992).

Another class of verbal measures that draw on the knowledge-base requires the participants to produce as many words as possible during a limited time interval, given
certain restrictions (e.g., words beginning with a certain letter or words belonging to a predefined semantic category). As regards the influence of age on such word fluency measures, some studies reported small age differences in these measures (e.g., Capitani, Laiacona, & Basso, 1998). However, several studies observed substantial age differences in such tasks (e.g. Parkin & Walter, 1992; Schaie, 1996), further suggesting that tasks that reflect rapid retrieval as well as knowledge are more likely to be negatively affected by aging effects than measures like vocabulary and fact recall.

Taken together, there is evidence that semantic memory is less susceptible to aging effects than episodic memory measures. Some studies showed no age-related variation. Measures that draw on the size of the knowledge base, rather than efficiency of retrieval from semantic memory, like vocabulary measures may even improve somewhat with age, at least until midlife, with some indications of a decline in old age. At the same time, measures that require rapid retrieval from semantic memory tend to be associated with age-related deficits. Hence, there is some heterogeneity within the domain of semantic tasks as regards aging patterns. Together with psychometric evidence (e.g., Lindenberger & Baltes, 1997) and findings that fluency performance may be selectively disrupted by head-injury (Haut, Petros, Frank, & Haut, 1991) and left frontal lesions (Benton, 1968) in the absence of effects on measures such as vocabulary, the possibility that knowledge and fluency measures load on separate semantic factors and that these sub-factors are associated with different age-related patterns in performance was examined further in Study III in this thesis.

Encoding and retrieval processes

Turning to accounts of the age-related deficits in declarative memory, focusing on the most age sensitive form, episodic memory, the cognitive processes associated with acquisition, or encoding, of to-be-remembered events, and those associated with retrieval of episodic information are essential. A longstanding issue is the extent to which the age-related deficits in episodic memory reflect inefficient encoding and retrieval processes.

Encoding

The levels of processing framework (Craik & Lockhart, 1972) highlighted the crucial role of encoding-related processes in accounting for subsequent success/failure in memory performance. More specifically, encoding processes were assumed to vary along a continuum from shallow to deep, where shallow refers to encoding processes directed towards superficial or physical aspects of the materials and deep encoding refers to processes directed towards the meaning of the event. If the word “apple” appeared in a study list, and you are instructed to count the number of syllables in that word (shallow tasks) you likely perform much worse in a subsequent test of recall than if you had been instructed to rate the pleasantness of the word (deep task). This is likely due to the fact that the latter task guides encoding towards meaning-based rather
than structural aspects of the to-be-remembered event, which results in a more durable record or trace of the event, which is beneficial in most episodic tests.

A number of findings suggest that both young and old adults adopt semantic processing to encode various kinds of materials (see Light, 1992; Kausler, 1994 for reviews). However, extended processing at the semantic level, called elaboration (Craik & Tulving, 1975), may create a memory representation that overlaps less with other representation in memory. A possibility is that whereas older adults pay attention to meaning-based aspects, they tend to encode the materials in a less elaborate or extensive, and more prototypical, fashion as compared to younger adults. A study by Rabinowitz and Ackerman (1982) was taken to support this hypothesis. The participants were presented with long lists of words from various categories (e.g., “horse” and “lion”) and were asked to generate a semantic associate to each word. At test, the generated associates served as cues to remember the nouns for half of the items. For the other half of the words, category labels (e.g., animal) were instead provided as cues to remember the nouns. Presumably due to the fact that young adult produced more specific associates, the young, but not the elderly, benefited from the associate cues. To take an example, younger adults would be more prone to produce the associates wild and domestic animal in response to “horse” and “lion”, respectively, rather than a semantic, but non-specific, associate such as “animal” for both species. Related findings by Rankin and Collins (1985, 1986) indicate an age-related deficit in the ability to generate elaborate associates, or cues, in response to words, as compared with young adults.

Apart from a difficulty to encode materials extensively at the conceptual level, there is evidence that older adults have problems to encode non-semantic contextual attributes of events relative to younger adults. Consider the following scenario: You heard that you should take sleeping pills to cope with your recent sleeping problems, or that the word GREEN was presented in the study list. Suppose that you later on are asked whether the recommendation to take pills was provided by your doctor or your friend Jack (who is a bank-teller), or whether the word GREEN was read in a male or female voice. In both cases you must distinguish between two potential sources of information. In other circumstances you may have to decide whether the remembered content information emanated from an internal or external source (e.g., did I hear this or imagined hearing it?). A large number of studies, using different contrasts among the sources (e.g. internal-internal, external-internal) as well as studies of memory for other non-content attributes (e.g., color), converge on the notion that older adults have problems of integrating and remember contextual aspects of studied events as compared with younger adults (Spencer & Raz, 1995).

The foregoing examples (i.e., semantic elaboration and encoding of perceptual features) concern encoding of attributes specific to the events (item-specific processing). It has been distinguished from processing of commonalities among studied events (relational processing; e.g., Hunt & McDaniel, 1993). Both types of processing are assumed to be beneficial to subsequent memory performance. From the examples above, and indications that older adults may processes relational information in a similar manner as younger adults, it is possible that age-related deficits in episodic encoding may be
characterized primarily by a failure to encode item-specific information. As judged from
the aforementioned examples this may include internally derived information (e.g.,
association, images) and externally derived (e.g., perceptual) information. Inefficient
encoding of perceptual or conceptual attributes will serve to produce memory
representations with larger overlap with other events in memory, in turn rendering them
more difficult to retrieve. Growing old may thus be characterized by increased
difficulties of encoding episodic events distinctively. We shall return to the possibility
that these problems may be overcome by implementations of specific instructions and
other forms of cognitive support.

Retrieval

The primary means to investigate the role of retrieval factors in memory research have
been to vary the extent of cue information available at time of retrieving events. In free
recall no such cues are available, in variants of cued recall some hint is provided (e.g. an
associate of a studied word or part of a studied sentence), and in recognition the to-be-
remembered item is re-presented. The principle line of reasoning is that if retrieval is the
primary locus of age-related differences in memory, providing substantial amounts of
cue-information, or re-presenting the information in form of a “copy cue”, the age
differences should be abolished. As was noted before, there is good evidence to believe
that recognition is associated with less pronounced age differences as compared with free
recall. From such results, the age-related problems of retrieval may be regarded to be
important in accounting for the age differences in episodic memory.

At the same time a substantial portion of the age-related variance in episodic
memory appears to remain when recognition is used as a means to assess memory,
indicating that only part of the age-related memory deficits is attributable to inefficient
retrieval processes. Also, whereas the expected pattern of magnified age differences in
free as compared with cued recall has sometimes been confirmed, some studies revealed
age differences of about the same magnitude in free recall and variants of cued recall
(Kausler, 1994).

At this point, it should be stressed that encoding and retrieval processes are highly
interdependent processes. That is, type of encoding determines what is stored, and hence
what will serve as effective cues at retrieval (generally: the greater overlap between
encoding and retrieval processes, the better).

Functional neuroimaging have provided a means to separate patterns of brain
activation related to encoding and retrieval processes. A consistent pattern across studies
using functional brain imaging techniques such as Positron Emission Tomography (PET)
and functional Magnetic Resistance Imaging (fMRI) is that episodic encoding is
associated with increased activation in the left frontal lobe. Retrieval of episodic memory
is by contrast associated with activation in right prefrontal regions (HERA, for
Hemispheric Encoding-Retrieval Asymmetry, see Cabeza & Nyberg, 2000 for a review).
The left frontal activation observed at time of encoding is usually taken to indicate
semantic processing. In fact, this brain activation is much the same as that observed when
participants are occupied with a semantic memory task. The right frontal activation associated with retrieval of episodes is assumed to reflect a “neurocognitive set”.

How are these patterns of brain activation affected by aging? Extant brain imaging studies involving age group comparisons seem to confirm that the neural correlates of encoding as well as retrieval processes are affected in old age. Specifically, regions in the left temporal lobe as well as in the left frontal lobe tend to be less activated by old as compared with young adults at time of encoding. At episodic retrieval, by contrast, older adults often show a bilateral instead of a unilateral (right) pattern of frontal activation (e.g. Cabeza & Nyberg, 2000), recently labelled HAROLD for Hemispheric Asymmetry Reduction in old adults, that has been regarded to be a result of compensatory processes or of dedifferentiation (for a review, see Cabeza, 2002).

In sum, there is evidence that advanced age is accompanied by some retrieval difficulties. However, even in situations when information is re-presented (recognition) substantial amounts of age-related variance remain, presumably due to encoding-related problems.

**R**e**c**ollective experience

In the recent decade there has been renewed interest in consciousness and recollective experience in memory research. A number of studies have been conducted from point of the hypothesis that retrieval in tests we have described as episodic may be associated with two distinct states of subjective awareness. To take examples of these states, recognizing a face in a crowd is often accompanied by retrieval of details such as where you have seen this face before. In other cases you may be absolutely certain that you have seen a face before, without being able to provide any information concerning the spatiotemporal circumstances of any past encounter with this person. You just know you have seen him/her before.

Following Tulving (1985b) and Gardiner (1988) a paradigm referred to as the Remember/Know (or R/K) paradigm has been used to examine these phenomenological aspects of retrieval. In this paradigm, typically used in conjunction with a yes/no recognition test, the participant is simply requested to respond “remember” if an item judged as old is accompanied by retrieval of any contextual aspect of the study event, as in the first example above. These contextual aspects include information of external sources as well as internally generated information (such as thoughts or images) from the original event. In case an item is recognized as old but does not bring to mind any contextual aspect of the original event, the participant is instead instructed to make a Know response.

A number of studies indicate that these states of awareness may be dissociated in the sense that some variables selectively increase the level of Remember responses without affecting the level of Know responses and vice versa. To take two examples, division of attention during encoding reduces the levels of Remember- but not Know responses whereas maintenance rehearsal increases Know- but not Remember responses (see Gardiner, 2001 for a review).
Several alternative theoretical views have been put forward to account for these and related findings. One of these suggests that remembering and knowing correspond to “autonoetic” and “noetic” consciousness that has been used to characterize retrieval from episodic and semantic memory. Thus Tulving (1985b) proposed that remembering and knowing reflect the output of the episodic and semantic memory systems, respectively. According to another theoretical view, remembering and knowing reflect differential involvement of recollection and familiarity (e.g., Jacoby, 1991). Recollection and familiarity are conceived of as independent processes. That is, in a particular memory situation, recollection, familiarity, or (presumably the typical case) both of processes contribute to memory performance. However, Remember and Know judgments are mutually exclusive in the paradigm described as the participant is required to make a Remember or Know judgment. As a consequence Know judgments must necessarily be restricted to familiarity in the absence of recollection. Provided that both familiarity and recollection contribute to Remember judgments, the rate of Know judgments will therefore underestimate the extent to which familiarity contribute to memory from the viewpoint of Jacoby’s dual-process theory (Yonelinas & Jacoby, 1995).

Of special interest in the present context is what variables promote higher levels of subsequent remembering. Quite consistent findings are that deep as compared with shallow levels of processing (Gardiner, Java, & Richardson-Klavehn, 1996), self-generation (Gardiner et al., 1996), and pictorial as compared with verbal presentation of information (Rajaram, 1996) selectively increase remembering. One way of conceptualizing these patterns is that variables that increase the distinctiveness of the to-be-remembered events promote remembering. Conversely, those processes that increase fluency, or familiarity-based processes, tend to increase knowing (Rajaram, 1996). As was noted earlier, one potential cause of older adults’ problem of remembering episodic events is that they encode less item-specific or distinctive information. Thus, one might expect diminished levels of remember experiences in old age.

Age differences in recollective experience

Is advancing age associated with a shift in the states of awareness that accompany retrieval? A data-base search identified a number of recognition studies relevant to this issue. A summary of the outcomes of 9 studies (with a total of 14 conditions) is presented in Table 1. This table contains information concerning the magnitude of the young-old difference (percentage of hits out of possible targets) obtained for Remember and Know judgments, respectively. The last column provides information concerning the nature of the encoding task employed (Guided versus Non-guided). We shall return to this distinction later.

As can be seen in Table 1, there are quite substantial differences between young and old groups in levels of Remember judgments in some conditions, while others were associated with relatively modest, or no, age-effects. The average across conditions is 15.9%. Thus, as far as the central tendency is concerned, old age appears to be associated with diminished levels of recollective experience. With regard to age differences in Know
judgments there is hardly any difference between young and old participants a judged from the average value (a 4.8 % difference consistent with marginally higher response levels in older adults); two studies indicate an age-related deficit (Perfect et al., 1995-2a) and most reported no significant difference. However, there is evidence of an age-related increase in frequency of Know judgments in the first five studies, as indicated by negative signs.

Table 1
Summary of studies involving age-group comparisons of Remember and Know judgments in recognition, including information concerning the mean young-old difference in terms of hit-rate (%) portioned into Remember and Know judgments, and type of encoding task (guided versus non-guided).

<table>
<thead>
<tr>
<th>Study</th>
<th>Young-Old Difference</th>
<th>Encoding Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Remember</td>
<td>Know</td>
</tr>
<tr>
<td>Fell (1992)</td>
<td>33</td>
<td>-36</td>
</tr>
<tr>
<td>Parkin &amp; Walter (1992)</td>
<td>32</td>
<td>-21</td>
</tr>
<tr>
<td>Parkin &amp; Walter (1992)</td>
<td>28</td>
<td>-15</td>
</tr>
<tr>
<td>Perfect et al. (1995)-1</td>
<td>35</td>
<td>-30</td>
</tr>
<tr>
<td>Perfect &amp; Dasgupta (1997)</td>
<td>23</td>
<td>-5</td>
</tr>
<tr>
<td>Perfect &amp; Dasgupta (1997)</td>
<td>30</td>
<td>-6</td>
</tr>
<tr>
<td>Norman &amp; Schacter (1997)</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Trott et al. (1999)</td>
<td>7</td>
<td>-1</td>
</tr>
<tr>
<td>Fell (1992)</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>Mark &amp; Rugg (1998)</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>Schaeter et al. (1997)-1</td>
<td>11</td>
<td>-1</td>
</tr>
<tr>
<td>Perfect et al. (1995)-2a</td>
<td>-1</td>
<td>21</td>
</tr>
<tr>
<td>Perfect et al. (1995)-2a</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Schaeter et al. (1997)-2</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Mean</td>
<td>15.9</td>
<td>-4.8</td>
</tr>
</tbody>
</table>

Note: Values with a negative sign indicate an age-related increase in responding.

In the foregoing section it was suggested that older adults’ tendencies to encode the materials in a less elaborate or distinctive manner may be expected to lead to diminished levels of Remember judgments. Specifically, remembering is per definition related to retrieval of contextual aspects of the original events, and as was noted before, older adults may have special difficulties when memory of contextual information is called for.

Age-related difference in response bias is an alternative account of the age-related differences. However, findings within this set of studies seem to be more in favor of an encoding-related failure on part of the elderly (e.g., Mäntylä, 1993). All of the aforementioned studies involved verbal materials. A study by Maylor (1995) used a different approach to study age differences in recollective experience. First, the materials were television theme tunes learned from an extra-experimental source. More crucially, the R/K states of awareness were inferred on basis of think-aloud protocols rather than by asking the participants to write down all the details concerning the source of the
memories. Importantly, the results converged with the main trends of the results in Table 1 by showing a negative age-effect in level of remember experiences whereas the level of answers indicative of know experiences were age invariant.

As is the case with studies of most areas of memory research, it should be noted that the evidence concerning age differences in recollective experience mostly rely on extreme age-group comparisons. In other words, we do not know whether the age-related differences in remembering are suggestive of a gradual decline from young to old age or not.

**Processing resource accounts**

In the previous sections we considered that specific aspects related to encoding (such as a inefficient encoding of item-specific/contextual information) and retrieval constitute likely bases of age-related deficits in episodic memory. A variety of specific explanations of age-related impairments in memory has been offered in analytically oriented studies that relate to the encoding/retrieval stages (e.g., suggesting that deficient imagery, elaboration, organization, and so forth is the locus of the age-related memory deficits). At the same time, there is much evidence that age-related variance is shared among different measures of memory that vary in the extent to which putative subcomponents or specific types of processing are involved. There is also much evidence that the age-related variance in memory, including episodic and semantic memory, is shared with measures of other cognitive abilities, measures often assumed to tap cognitively and neuropsychologically separate functions (e.g., Salthouse, Fristoe, & Rhee, 1996). A possibility is that these patterns of results are attributable to the fact that a substantial portion of the age-related deficits in memory (and other cognitive abilities) are mediated by reductions in the quantity of some general-purpose capacity or “mental resource”. As this processing resource view has been a major driving force in cognitive aging research during the last decade or so, we shall consider some of its variants. Salthouse (1988) distinguished between three resource: processing speed (e.g., Salthouse, 1991, 1996), working memory capacity (e.g., Salthouse & Meinz, 1995), and inhibition (e.g., Earles et al., 1997). Next, a brief outline of variants of the processing resource view will be provided, including speed, working memory and a third resource that has been increasingly discussed as a potential mediator of age-related changes in memory, that may be regarded to include inhibition, namely executive functions.

**Processing speed**

There is abundant evidence of an age-related slowing across a variety of cognitive and perceptual measures. These include simple reaction time, choice reaction time, memory scanning and measures of perceptual speed as reflected by for example Digit-Symbol test of the WAIS battery (Wechsler, 1981). The age-related slowing is also evident in ERP components (e.g., P300; Bashore, Osman, & Heffley, 1989).

In the version of a speed theory put forward by Salthouse (1996), two speed-
mechanisms were proposed to account for the negative age differences observed in a variety of cognitive tasks: (i) a limited-time mechanism, and (ii) a simultaneity mechanism. The first mechanism concerns the time taken to complete a given mental operation. Specifically, the time available for subsequent mental operations may be too slow in old age to be completed within the time-frame available. If for example the task is to generate associations, and if the rate of processing is lowered with advancing age, a smaller number of associations will be generated within a given time frame. If older adults suffered from this aspect of slowing only, one might predict that given additional time, for example to encode the to-be-remembered items in a word-list experiment, the elderly should be able to catch up with the younger adults, that is, compensate for their lower rate of processing. As already noted, there is, however, little support for the idea that conditions with long as opposed to short study time serve to eliminate (or even reduce) age differences in memory (Verhaeghen et al., 1993). However, the second of the proposed mechanisms (simultaneity) departs from the notion that processing of a mental operation is contingent on the products of previous operations. The latter Salthouse (1996) argued, may be lost if speed of processing are low. Thus, to the extent that success on the task at hand is dependent on the joint processing, or integration, of a number of units of information, older adults will suffer from slower rate of processing. In other words, rate of processing may restrict the quality of processing that may be achieved, not only the time taken to complete given processes.

In most studies used to investigate the relation between processing speed and memory, some composite of measures used to assess perceptual speed (e.g., digit-symbol) age; Salthouse, 1996). Hierarchical regression analysis, in which speed measures are entered before and after age have generally shown that the amount of the age-related variance in memory performance is reduced considerably, often by more than 60 percent, when individual differences in speed are partialled out (Salthouse, 1993a). This has been taken to indicate that a considerable proportion of the age-related deficits are attributable to age-related reductions in speed of processing.

Working memory

Working memory is required to simultaneously process and store new information. As such, this capacity is necessary for the successful completion of a variety of higher-order cognitive tasks, including comprehension of language, reasoning, and transfer of information to long-term memory. In the model by Baddeley (1992) the working memory system consists of two slave system devoted to storage of auditory and visuospatial information and a central executive component, coordinating storage and processing of the available information.

Measures of working memory that have been adopted in cognitive aging research include loaded reading and digit span. At the same time as the to-be-memorized items are held in memory, the participant has to perform some additional task on the same materials (e.g., judge whether the sentences are true/false). Such complex working memory tasks
are associated with quite sizeable age differences (Salthouse, 1991; Kausler, 1994 for reviews) in contrast with traditional short-term tasks such as digit-span, that primarily measure storage capacity (rather than processing). Also, in complex tasks, processing capacity, in particular, is negatively affected by the aging process (e.g., Salthouse & Babcock, 1991).

In a similar manner as measures of perceptual speed, researchers have used measures of working memory as a means to estimate the relative amount of the age-related variance in other cognitive measures that is predicted by the working memory measures. In general, these results show that differences in working memory capacity account for a considerable portion of age differences in episodic memory.

The interrelations among measures of speed, working memory, and episodic memory, have been investigated in several studies. Several studies have suggested that speed is a driving force behind working memory deficits in old age, although measures of working memory often appear to contribute to variance in memory over and above speed (Verhaeghen & Salthouse, 1997). Together these factors may thus account for a large share of the age-related variance in memory (e.g., Park, 2002). A study by Park et al. (1996) allowed for a comparison of the role of the two resources across memory tests expected to vary with regard to degree of resource demands: free recall, cued recall, and recognition. Interestingly, the results indicated that with regard to performance on the more demanding test (free and cued recall), both speed and working memory contributed to age effects in episodic memory, whereas speed alone accounted for a majority of the age-related variance in recognition memory. Finally, when a memory factor comprised by all of the tests was considered, the results indicated the central role of speed. That is, a model in which this factor was hypothesized to mediate working memory capacity, which in turn mediated episodic memory performance, was found to fit the data well. Thus, there is support of both the speed and working memory versions of the resource reduction account of age differences in memory, but also indications of the need to consider a dual resource account.

Executive functions

In recent years, a third type of account has emerged. In contrast with the foregoing resource accounts, this account was driven by neuropsychological findings that indicated the crucial need to consider a class of functions associated with the frontal lobe regions of the brain, functions that are believed to subserve “conscious, strategic, goal directed cognitive activity” (Luszcz & Bryan, 1999, cf. West, 1996). At the neural (structural) level, there is evidence that age-related deterioration of the neural substrate are especially large in prefrontal cortices, even though aging appear to have rather widespread effects on the brain as judged from morphological measures (Raz, 2000). As regards functional evidence, there are as we have noted before results showing that activity in prefrontal regions is associated with poorer memory performance in older adults.

Regarded as a resource, the executive functioning account differ from speed and working memory in that it is often conceived of as a collection of functions rather than a
single factor (although researchers have often used the term executive function or frontal lobe function as if it were a unitary construct). It should be noted that there is an obvious connection between executive functions and working memory. Specifically, executive functions are related to the central executive that may be challenged in more demanding working memory tasks. At the behavioral level a wide variety of measures have been used as indicators of executive functions, including measures derived from the Wisconsin card sorting test, planning tasks like Tower of Hanoi (Rönnlund, Lövdén, & Nilsson, 2001), the Stroop-test, and measures of word fluency (that were previously classified as semantic memory measures) to name a few.

In the same manner as with measures of speed and working memory, regression analyses have been used as a tool to investigate the possibility that executive function mediate age differences in episodic memory. The results are mixed. Specifically, in some studies the age-related variance in memory was reduced or substantially when performance on the executive function measures were partialed out (Bryan et al., 1999; Troyer, Graves, & Cullum, 1994), and in other cases only a weak or nonexistent link was found between age, memory and measures of executive functions (Spencer and Raz, 1994).

Of special interest here, two studies have adopted executive function measures to investigate the hypothesis that the patterns of lower rates of remember reports with in studies is the result of diminished executive functions in old age. This hypothesis is based on indices from a variety of studies showing that (a) there are age-related deficits in frontal/executive measures (b) the elderly have particular difficulties to remember contextual details which, per definition, are required to make a valid R-judgment, and (c) evidence that the frontal lobe is critical to the ability to remember contextual information. Parkin and Walter (1992) considered measures derived from three tests often referred to as executive functioning tests (labelled “frontal lobe tasks”). The first was FAS, which requires the participants to produce as many words beginning with F, A, and S, respectively during one minute. The second was WCST (four measures) and the third test was the Embedded Figures Test. Some of these measures were related to remember experiences in old adults. However, this held true within a sub-group of the total sample. Perfect and Dasgupta (1997) found a relative lack of relation between age, recollective experience, and executive function measures.

Even though impairments in executive or “prefrontal lobe functioning” seem to be a dominant explanation of age-related differences in memory among neuroscientifically oriented researchers (see Raz, 2000), attempts to predict the age-related variance in memory from measures assumed to reflect executive functions have, as already noted, been somewhat mixed (e.g., Spencer & Raz, 1994). A potential reason for the relative lack of consistency as regards the association between memory and executive functions may be that the measures reflect quite different aspects of this construct and that some measures have insufficient reliability (e.g. Rabbitt, 1997).

Taken together, the evidence regarding the relation between age-related deficits in memory and measures of variants of the processing resource account suggests that individual differences in measures of executive functioning, and, perhaps more
consistently, working memory and speed, predict a considerable portion of the age-related variance in memory. Here the focus was on episodic measures, but at least as regards perceptual speed there is evidence that most of the age-related variance in semantic memory in old age is predictable from perceptual speed measures (e.g., Lindenberger, Mayr, & Kliegl, 1993). However, as noted by Luszcz and Bryan (1999) the aforementioned resources may represent partly overlapping constructs. Specifically, the simultaneity mechanism proposed by Salthouse (1996) seem to presuppose a working memory and executive functioning is a component is an important component in Baddeley’s (1992) working memory model.

Also, a high degree of co-variation among age differences in memory measures and differences in markers of the resource constructs does not necessarily imply a causal relationship. In fact, is a growing body of findings that suggest the possible existence of a higher-order age sensitive “common factor”, that is, a factor that is more inclusive than is suggested by any of the aforementioned processing as measures of the hypothesized mediating factors (e.g., speed) may also be indicators of the common factor (e.g., Salthouse, 2001).

Interesting in this regard are result suggesting that substantial amounts of the age-related variance in measures of memory are shared with measures of sensory functioning and other non-cognitive factors like grip-strength, possibly suggesting that cognitive as well as noncognitive factors draw on a common age-sensitive factor (e.g., Anstey, 1999; Baltes & Lindenberger, 1997; Salthouse, Hancock, Meinz, & Hambrick, 1996).

Finally, any complete mediator or common cause account ultimately need to specify the neurological underpinnings of the positive manifold phenomenon among age-related deficits. At this point, research relating age-related losses in memory functioning to changes in cerebral white matter in old age (e.g., loss of myelin; Gunning-Dixon & Raz, 2000) and research examining the relation between age-related deficits in of dopamine functioning, memory, and other cognitive abilities (Bäckman et al., 2000), represent examples of recent trends.

**Encoding support and modifiability of age-related memory deficits**

As was discussed previously, there is good reason to believe that portions of the negative age differences in episodic memory measures stem from inefficient encoding processes. One possibility is that older adults do not spontaneously engage in effective encoding processes, but are able to do so given that some kind of support in the form of instructions, hints, or variations in the materials is provided, a view that will be referred to as a production deficiency view (Salthouse, 1991). A related possibility is that older adults do not process information at encoding in an optimal way due to limitations in basic cognitive or neural resources (Logan et al., 2000) but that these limitations are overcome given that cognitive support is provided.

In the framework by Bäckman (1985a, 1989), two superordinate concepts were proposed to account for aging patterns in episodic memory: spontaneous recoding and compensation. Recoding was (cf. Tulving, 1983) used to refer to processes or operations
that yield richer or more elaborate representations of the initially registered information. Imagery, organization, and cross-modal transformation were provided as examples. Negative age differences in episodic remembering, Bäckman argued, are attributable to age-related problems of carrying out such processes spontaneously. On the other hand, older adults may be hypothesized to compensate for these deficits by provision of external or internal support (Bäckman, 1985, 1989; Herlitz, Bäckman, & Lipinska, 1990). External support has been used to refer to guidance in the form of instructions or cues provided by the experimenter, whereas internal support has been used to refer to quantitative or qualitative alterations in the nature of the materials that enhance subsequent retention. Presenting concrete as opposed to abstract materials, presenting items in an organized rather than unorganized format, or at a slower rate than in the baseline condition, were provided as examples of internal support.

A similar view was put forward by Craik (1983, 1986). In his terms, age differences in memory performance is hypothesized to be a function of the extent to which a given memory task requires self-initiated processes on the one hand and the amount of environmental support available on the other. The requirement of self-initiated processes, he assumed, should decrease with increasing levels of environmental support. Hence, deficiencies in carrying out self-initiated processes on part of the older adults may be offset by provision of environmental support.

The literature on utilization of cognitive support in young and old adults is extensive, and the intent is not to give a comprehensive review here (see Bäckman, Mäntylä, & Herlitz, 1990; Craik & Jennings, 1992; Kausler, 1994). Rather the results pertaining to a few encoding manipulations assumed to guide encoding towards further processing of conceptual or perceptual detail (rather than relational information) will be given.

Internal and external support: A few examples of results

Studies consistently show an advantage of self-generated over experimenter-provided materials (the generation effect; Slamecka & Graf, 1978). As judged from the results of several studies both young and old adults improve following instructions to generate rather than read or hear words, and the magnitude of age differences is similar for provided and self-generated words (e.g. Dick, Kean & Sands, 1989; Johnson, Schmitt & Pietrukowicz, 1989; McFarland, Warren & Crockard, 1985). In a similar vein, instructions to visually imagine verbal materials typically boost subsequent memory performance, possibly related to the activation of a visual code in addition to a verbal code (Paivio, 1971). However, age differences following provision of imagery instructions are often similar, or somewhat larger, as compared with the control condition in which no instructions are provided (e.g. Bäckman & Nilsson, 1984; Dirkx & Craik, 1992; McCauley, Eskes, & Moscovitch, 1996).

In a related manner, presenting words in two modalities rather than one should serve to add a sensory attribute to the representations, and thus increase the number of cues on which retrieval may be based. Results by Arenberg (1967) indicated that age differences
were reduced following bimodal as compared with unimodal presentation (see also Bäckman, 1986), in line with a compensation view. However, in a comparable study by Stine, Wingfield, and Myers (1990) that concerned memory for television news studied under each of three conditions: listen, listen plus read, and listen plus watch (i.e. with visual track), the opposite pattern was observed. That is, only the younger participants were able to utilize the bimodal cues to improve memory as judged from subsequent recall of the news.

Another effect tied to task properties is the picture superiority effect. It refers to the common finding of better retention of pictures as compared to the corresponding words (Paivio & Csapo, 1973), which as noted before, may be due to the fact that availability of perceptual attributes in the case of pictures allows for encoding of an increased number of attributes on which retrieval may be based. This effect has typically been found to be similar in old and young adults, or to be somewhat larger for young (Park, Puglisi & Sovacool, 1983; Rissenberg & Glanzer, 1986; Maisto and Queen, 1992).

As judged from findings such as these the viability seem low of a simple model predicting reduced age differences in episodic memory following provision of support in form of instructions or materials that promote encoding of item-specific information. The aforementioned examples of findings rather suggest that the magnitude of age different remains invariant across levels support exacerbated in the more supportive conditions.

As was clear from the review of prior studies using the R/K method (Table 1), the tendency of data is that of lower frequencies of remember experiences in groups of older as compared with younger adults. On basis of part of the studies in Table 1 Perfect et al. (1995) proposed the following hypothesis: In conditions were older adults are left to their own device (i.e., non-guided encoding conditions), they are less likely to engage in extended processing of the materials, but tend to engage in rote or maintenance rehearsal. As mentioned earlier, this might be expected to increase levels of K-judgments, and yield lower levels of R-judgments. Provided guidance (e.g., in form of specific instructions) old adults may on the other hand shift from such a strategy and engage in more effective encoding processes. Thus, in accord with this hypothesis, age differences in levels of R- and K-judgments should be minimal following provision of encoding support.

The data in Table 1 appear to offer some support for this hypothesis. Specifically, the mean age difference in R-judgments in conditions that are guided by some orienting task is 5.7 %, whereas the corresponding value is 23.6 % across the no-guided conditions. In addition, there is evidence of a reversed trend for K-judgments. Here, the average young-old difference is 7 % across the guided conditions and −13.8 % across the non-guided conditions (i.e., consistent with an age-related increase in knowing).

These patterns of results are largely based on comparison between studies. Thus, the differences in outcome may reflect other variations than presence or absence of guidance (i.e., sample composition, materials and so forth). Nevertheless, the data seem to offer some support for the idea that age-related impairments in distinctive encoding may be “repaired” by provision of environmental support, in line with a production deficiency hypothesis. However, as regards item recall (overall recognition
and measures of recall) the dominant picture across studies seem to be that younger adults benefit from various kinds of encoding support at least as much as the older adults.

How should the latter type of result be interpreted? From point of a production deficiency view, it would appear to be the case that the young adults do not always employ optimal encoding strategies either. Alternatively, the assumption that increases in environmental support necessarily reduces the need to carry out effortful or self-initiated processes may be questioned. Certain types of encoding support possible guide processing without altering the demands on critical resources (e.g., self-initiated processing). To avoid circular reasoning, independent measures of the construct (e.g. resource) hypothesized to mediate the age differences should be adopted. This has rarely been done in studies of encoding support. Next, enactment, the specific form of encoding support examined as part of the present thesis (Study I and II) will be introduced.

**Enactment**

In episodic measures, whether verbal (word-list recall or recognition) or nonverbal (face recognition), the encoding phase is passive in the sense that no overt activity is involved. By contrast, actions constitute an important class of events in every day memory. The neglect of studies of memory of enacted events inspired Cohen (1981) to develop a paradigm, referred to as the subject-performed tasks, or SPTs for short. In its basic form, this paradigm requires the participants to perform a series of one-step actions in response to verb-noun phrases such as "lift your hat", "scratch your nose", or "throw the ball". In the control condition participants read or listen to these phrases without performing the actions (cf. Wolff & Levin, 1972).

Independently, Engelkamp and Krummacker (1980) and Saltz and Donnenwerth-Nolan (1981) used a virtually identical procedure. These researchers had a slightly different focus in that they were primarily concerned with how motor encoding affects subsequent retention of verbal materials. That is, enacted encoding was seen as a form of orienting task rather than a separate class of events or materials (cf. Nyberg, 1993, 1995).

It is a well-established finding that enactment enhances retention of phrases. The superior levels of recall of enacted over non-enacted phrases, often in the range of a 20 to 30 percent advantage, is known as the enactment effect. It is observed in free recall (e.g. Bäckman & Nilsson, 1986; Cohen, 1981; Engelkamp & Krumnacker, 1980; Kormi-Nouri, 1995) cued recall (e.g. Saltz & Donnenwerth-Nolan, 1981) and recognition (Engelkamp & Krumnacker, 1980; Nilsson & Craik, 1990).

Theoretical accounts of the enactment effect, such as those put forward by Bäckman and Nilsson (1984) and Engelkamp and Zimmer (1984) have emphasized that execution of the actions provides additional sensory and motor attributes (see Engelkamp 1998 for a review). Bäckman and Nilsson (1984, 1985; Nilsson & Bäckman, 1991) noted that enactment of phrase events requires recruitment of additional sensory systems (haptic, olfactory, gustatory) as compared with nonenacted items (“smell the flower”, “eat the biscuit” et cetera), which are nominally present in one or two modalities (read/heard). In addition, visual and tactual exploration of
features related to the physical appearance of the objects (e.g. color, texture) may serve to enrich the encoding of the events. In line with the latter hypothesis Bäckman and Nilsson (1990) demonstrated build up and release of proactive interference (Wickens, 1970) along physical dimensions such as weight, color and size of the concrete objects that were used. However, a substantial enactment effect is observed also when participants perform the actions with imaginary objects. Also, the effect is observed for actions that do not require the provision of an external object (Engelkamp, 1998).

The account put forward by Engelkamp and his co-workers (Engelkamp, 1998; Engelkamp & Zimmer, 1984; Zimmer & Engelkamp, 1989) emphasized the role of motor aspects in explaining the superior retention of enacted materials. Inspired by Paivio’s (1971) dual coding theory, that postulates verbal and visual-imaginal codes (or systems), they hypothesized that enacted materials may benefit from the activation of an additional motor code.

Several lines of evidence indicate that motor information is critical for the enactment effect. First, enactment typically improves recall and recognition as compared with observing a model perform the action (e.g. Arar, Nilsson & Molander, 1993; Conway & Dewhurst, 1995; Dick & Kean, 1989; Engelkamp & Krummacker, 1980; but see Cohen, 1981, 1983). Second, instructions to enact are usually more beneficial to subsequent memory performance than instructions to imagine oneself performing the actions (Conway & Dewhurst, 1995; Engelkamp & Zimmer, 1984; Guttentag & Hunt, 1988; but see Helstrup, 1987). These findings indicate that the enactment effect is not reducible to visual imagery, processing of dynamic visual information, or planning. However, these and related findings (Koriat, Ben-Zur & Nussbaum, 1990; Goschke & Kuhl, 1993) probably indicate that such information contributes to the difference between verbal and enacted encoding.

Second, memory of enacted materials is selectively impaired when interpolated activity between study and test is motoric rather than kinematic (Engelkamp & Zimmer, 1985) or verbal in nature (Cohen, 1989; Kausler, Wiley & Lieberwitz, 1992; Saltz & Donnenwerth-Nolan, 1981). Also, instructions to re-enact at time of test may enhance memory of actions (Engelkamp, Zimmer, Mohr, & Sellen, 1994; but see Kormi-Nouri, Nyberg, & Nilsson, 1994), presumably due to the fact that the representation of the encoded motor information is reactivated at time of retrieval (cf. Tulving & Thompson, 1973).

Finally, two PET-studies (Nilsson et al., 2000; Nyberg et al., 2001; cf. also Heil et al., 1999) provided results consistent with the view that motor information is critical to understand the enactment effect. At verbal retrieval of initially encoded actions the primary motor areas were re-activated. No such activation was present in the verbal condition. Interestingly retrieval of actions that had only been imagined was associated with activation of motor areas as well, but to a lesser extent. The presence of a motor component gives enactment a special status in the domain of support types. It could thus be that it has capacity of altering the magnitude of age-related differences in episodic measures in a significant fashion, an issue that will be considered next.
Prior aging studies of enacted materials

A study by Bäckman and Nilsson (1984) was the first aging study that examined the possibility the magnitude of age differences in memory is altered following enactment. A group of young adults in their 20s were compared with middle-aged subjects in their 40s and older adults in their 60s in free recall of action commands that were either acted out or studied verbally (bimodally). A third encoding condition included single nouns. In recall of sentences and nouns, pronounced negative age differences were observed. By contrast, recall of enacted sentences was associated with non-significant age differences, a result replicated in two follow-up studies (Bäckman, 1985b; Bäckman & Nilsson, 1985). These findings were taken to support the viability of the compensation framework discussed earlier. Specifically, the presence of multimodal cues and availability of contextual features were hypothesized enable the elderly participants to overcome problems of self-initiated processing. These indications of a possible exception to the general pattern of negative age differences in episodic memory sparked much interest.

To date, a number of age-group comparisons on memory for enacted materials have been published. A summary of the results from these studies is provided in Table 2. The studies are grouped on basis of the principle outcome (last column), that is according to whether enacted encoding reduced the negative age differences observed in the control condition (Yes) or not (No). In two studies reduced age differences were observed/not observed depending on the level of a third variable. The mean age difference in percentage items recalled is furthermore provided (In cases were the studies included multiple experiments/ conditions these values represent the weighted mean).

As can be seen, the original finding of attenuated age differences in recall of SPTs as compared to nonenacted control tasks (e.g. words or sentences) has been replicated in a few subsequent studies (Brooks & Gardiner, 1994, Dick & Kean, 1989; Nyberg, Nilsson & Bäckman, 1992). However, all of the studies reveal some advantage on part of the younger group. More importantly, the majority of studies seem to indicate comparable following enacted and nonenacted encoding. In fact, the weighted averages across all studies reveal only very modest difference with regard to the magnitude of the young-old difference in enacted (M = 15.9) and nonenacted conditions (M = 19.5).

Discrepancies among studies may of course depend on some third variable, perhaps a procedural factor. The search for such a moderator variable has generally been unsuccessful in the sense that age differences have been found to generalize across third variables. This pertains to several studies in the No-reduction category.

To provide a few examples, results by Knopf (1992) indicated that both young and old performed better when concrete objects were provided at encoding rather than when the phrases were enacted with imaginary objects. However, this effect was seen regardless of age group. Also, familiarity of actions does not appear to produce a difference with regard to age effects for enacted information Knopf (1991). Other encoding manipulation used in conjunction with enactment has also revealed little evidence that the level of a third affects the basic outcome, including whether the actions were self-generated or provided by the experimenter (Lichty, Bressie, & Krell, 1988).
<table>
<thead>
<tr>
<th>Study</th>
<th>Mean Y-O differencea</th>
<th>N</th>
<th>Reduction?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bäckman &amp; Nilsson (1984)</td>
<td>11</td>
<td>72</td>
<td>Yes</td>
</tr>
<tr>
<td>Bäckman &amp; Nilsson (1985)</td>
<td>6</td>
<td>72</td>
<td>Yes</td>
</tr>
<tr>
<td>Bäckman (1985b)</td>
<td>1</td>
<td>40</td>
<td>Yes</td>
</tr>
<tr>
<td>Dick &amp; Kean (1989)</td>
<td>8</td>
<td>96</td>
<td>Yes</td>
</tr>
<tr>
<td>Nyberg et al. (1992)</td>
<td>15</td>
<td>72</td>
<td>Yes</td>
</tr>
<tr>
<td>Brooks &amp; Gardiner (1994)</td>
<td>6</td>
<td>72</td>
<td>Yes/No</td>
</tr>
<tr>
<td>McDonald-M. et al. (1996)</td>
<td>20</td>
<td>72</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Earles et al. (1999)</td>
<td>21</td>
<td>96</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Nilsson &amp; Craik (1990)</td>
<td>24</td>
<td>64</td>
<td>No</td>
</tr>
<tr>
<td>Knopf (1992)</td>
<td>20</td>
<td>200</td>
<td>No</td>
</tr>
<tr>
<td>Cohen et al. (1987)</td>
<td>12</td>
<td>54</td>
<td>No</td>
</tr>
<tr>
<td>Knopf &amp; Neidhardt (1989)</td>
<td>17</td>
<td>60</td>
<td>No</td>
</tr>
<tr>
<td>Earles (1996)</td>
<td>20</td>
<td>203</td>
<td>No</td>
</tr>
<tr>
<td>Guttentag &amp; Hunt (1989)</td>
<td>21</td>
<td>32</td>
<td>No</td>
</tr>
<tr>
<td>Norris &amp; West (1991)</td>
<td>18</td>
<td>80</td>
<td>No</td>
</tr>
<tr>
<td>Knopf (1995)</td>
<td>16</td>
<td>48</td>
<td>No</td>
</tr>
<tr>
<td>Lichty et al. (1988)</td>
<td>13</td>
<td>24</td>
<td>-</td>
</tr>
<tr>
<td>Kausler et al. (1990)</td>
<td>13</td>
<td>24</td>
<td>-</td>
</tr>
<tr>
<td>Phillips &amp; Kausler (1992)</td>
<td>23</td>
<td>96</td>
<td>-</td>
</tr>
<tr>
<td>Kausler et al. (1992)</td>
<td>13</td>
<td>144</td>
<td>-</td>
</tr>
<tr>
<td>Wiley &amp; Kausler (1993)</td>
<td>11</td>
<td>64</td>
<td>-</td>
</tr>
<tr>
<td>Mean (weighted)</td>
<td>15.9</td>
<td>19.4</td>
<td></td>
</tr>
</tbody>
</table>

Note. aweighted average across conditions/experiments.

Overall task difficulty may be a clue to the reason for the deviant findings (Engelkamp & Cohen, 1991, cf. Salthouse, 1991). Specifically, most studies that found reduced age differences in the enacted condition used short lists (12 items; e.g., Bäckman & Nilsson, 1984). Also, most of these studies averaged the results across multiple trials. A three-way interaction, involving age, encoding, and list-length (Brooks and Gardiner, 1994) and study trials (McDonald-Mizczak et al., 1996), was observed in the studies within the Yes/No-category. These effects reflected a selective reduction of the young-old difference in the enacted condition provided that the lists were short, and following repeated trials, respectively. It could thus be that the younger adults leveled off at near functional ceiling levels, while the elderly improved somewhat across trials in some of the studies within the Yes-category. One could also note that in the original studies encoding served as a between-subjects variable. In addition, in the study by Nilsson and Bäckman...
(1984), the age difference in the nonenacted condition is the largest across studies. Thus, it possible that the interaction in this case in part reflects an unusually high performance level in the young nonenactment group. Of direct relevance, a follow-up study (Nyberg et al., 1992), in which the same materials was used as in the original study, showed a pattern of more modest reduction of the age differences in the enacted condition. In fact, several of the studies in the Yes-category report a surprising lack of enactment effect in the young group, something which appears to restrict the generality of these results (cf. also Knopf, 1995). Thus, from point of several observations an outcome consistent with age-related deficits may be expected to show up in recall of enacted materials. These patterns are good agreement with the results of another paradigm known as “activity memory”. In this case, the participants take a series of longer tests requiring overt/covert activity rather than perform discrete actions. Quite consistently, verbal recall of the activities performed is associated with negative age differences (Bromley, 1958; Kausler, 1994 for a review) in line with findings of substantial age effects in recall of enacted sentences.

Relatively few aging studies have examined memory of enacted materials under more supportive conditions (e.g., recognition). In this context it is interesting to note that there are some indications that older adults are able to compensate for episodic deficits given that the encoding support is provided in conjunction with substantial retrieval support. In a study by Bäckman and Larsson (1992) young, middle-aged and old participants encoded nouns under two different conditions: without or with real objects. Memory for these were tested by free recall or category-cued recall. The results were taken to suggest that older adults are able to compensate for part of the age-related deficits, provided that both encoding and retrieval is guided; when concrete objects were present during encoding and cues were provided at retrieval age differences were reduced, not otherwise. In a related vein, AD (Alzheimer’s Disease) patients appear to be benefit from enacted support, but only provided that cues are provided at retrieval (Herlitz, Adolfsson, Bäckman & Nilsson, 1991; Karlsson, Bäckman, Herlitz, Nilsson, Winblad & Österlind, 1989). However, Craik and Nilsson (1990) failed to detect an interaction between age, encoding type and test, using free versus cued recall (noun cue) as the dependent measures, indicating that age differences following enactment may persist even though environmental support is provided at test.

As regards recognition, the results are mixed. Knopf and Neidhardt (1989) observed similar age effects for enacted and nonenacted sentences in free recall, but in a subsequent test of recognition of the same pool of commands the age differences were eliminated in the enacted condition. However, as is often the case with recognition of enacted materials, the hit rate for the young group was at a high level (> 90%). The interaction may thus have reflected a ceiling effect for the young (see also Nilsson & Craik, 1990). Knopf (1991) used a longer retention interval to lower hit-rate. In this case, the magnitude of the negative age differences was similar for enacted and nonenacted items.

From the specific production deficiency hypothesis put forward by Perfect et al. (1995), a pattern of reduced negative age differences in levels of remember experiences following enacted encoding might be predicted even in the absence of an age by encoding interaction in overall recognition. This is so because the age-related increase in rote
rehearsal believed to underlie the pattern of lowered levels of Remember judgements and
higher levels of Know judgements (as observed in some studies), should be prevented by
offering an orienting task. In other words, following enactment a much higher proportion
of recognized items might be remembered than known especially so in groups of elderly
participants.

In sum, the initial observation that enactment may serve to reduce some, if not all, of
the negative age differences observed in recall has been difficult to replicate. Rather, the
overall patterns of results suggest that there are reliable age-differences in recall of
enacted materials, and that these approach those in the control conditions, suggesting that
young and old benefit from this type of support to a similar extent (cf. Craik & Jennings,
1992; Kausler, 1994). Some studies indicate that encoding support paired with retrieval
support retrieval may allow the elderly to compensate, however. The results pertaining to
recognition of enacted information is sparse and less conclusive than for recall. From the
production deficiency hypothesis by Perfect et al. (1995) portioning of recognition into
components of recollective experience might be expected to yield an age by encoding
interaction consistent with selectively spared levels of remembering following enactment
in groups of older adults, a possibility that was examined in Study II in the present thesis.
Finally, large-scale investigations, covering a broader range of the adult life-span is
missing in the literature.

Age versus time- and cohort-related influences

Chronological age is an index of the processes associated with maturation. Up to this
point we have considered data from studies wherein this index was used to infer changes
in groups of individuals, by assuming that average differences between one or several
groups differing in age established at a specific point in time provides a fair estimate of
average maturational change.

However, as was noted by Schaie (1965), a more comprehensive model of
developmental changes, must probably consider that chronological age is not the only
possible source of influence on cognition/behavior. Specifically, two additional factors,
time and cohort were suggested to have potential influence. The time factor has been used
to denote a set of history-bound events that potentially affect the observed behavior of an
individual or a group of individuals across consecutive measurement intervals (e.g., a
shift in educational practice that make people more familiar with the kinds of tests used).
Cohort effects reflect common generation-bound conditions, or more generally, effects
shared by common time of entry into a system. The usual manner is to operationally
define the latter in terms of entry into life. In other words, birth-year is often used as an
index of cohort-related influences.

Together, Schaie argued, these factors (age, time, and cohort) make a joint
contribution to a given “response”, or observed level of functioning of the organism at a
given time point. Of primary concern, two out of the three factors are necessarily
confounded using the designs (cross-sectional and longitudinal) typically adopted by
developmental researchers. By use of a more extensive design, and analyses allowing that
two out of the three proposed factors (age, time, and cohort) vary, Schaie argued that it should be possible to decompose the effects associated with each factor.

Rather than going into the complex issues at the moment (some additional features of the proposed design will be spoken of later in connection with a presentation of the overall design of the Betula study, we shall focus on cohort factors and cross-sectional versus longitudinal aging patterns on measures of episodic memory in particular. Specifically, all of the studies cited earlier concerned cross-sectional age differences, and a major threat to the internal validity of these data are cohort factors. Changes observed across time are devoid of cohort-related influences, and thus improve upon the cross-sectional design in one important aspect.

**Cohort factors**

People differing in age tested at a single point of time differ not only in number of years elapsed since date of birth; they were born and grew up during different conditions. These generation-specific conditions likely differed along a number of dimensions. Suppose that with historical time, nutrition, or quantity of education improved, and that these factors served to yield a positive effect on average levels of cognitive performance. In such a case the earlier born cohorts might be at disadvantage later, not exclusively due to the fact that they are older, but because they grew up during less advantageous conditions.

As already noted, a very high percentage of the published studies on aging and memory are cross-sectional and are hence susceptible to potential cohort effects. Some researchers have defended the predominant use of cross-sectional memory studies by assertions that cohort effects are unlikely to influence the performance on the kind of laboratory tasks used to assess episodic memory (e.g., Kausler, 1991). Such a claim possibly reflects the belief that episodic memory functioning, in line with other fluid aspects of cognition is driven by biological factors (i.e., unlike measures of semantic memory/crystallized measures; cf., Gf-Gc theory; Horn & Hofer, 1992).

However, quite substantial time-lag gains have been observed for measures proposed to reflect fluid aspects of cognition during the 20th century. To provide an example, men have long been tested as part of enrollment in military service. Across successive generations the scores on the same tests taken at the same age have improved (Flynn, 1987; Neisser, 1998). If this is an effect of cohort factors, such that the peak level attained in young adulthood increased across successive birth cohorts (and no time-related improvements occurred later) cross-sectional age comparisons may be seriously distorted (i.e., negative age differences would, partly or wholly, reflect cohort differences).

Education may be an important cohort factor. Part of the commonly observed association between cognition and education may result from selection factors, that is, more cognitively able individuals are more likely to attain higher levels of education. However, as judged from a large body of evidence, it appears to be that education has a direct influence on a variety of cognitive tasks (Ceci, 1991). To provide examples of the evidence, children born at the beginning of the year tend to outperform those born at the end of the year in cases where they attend to different grades (despite the fact that they...
differ minimally with regard to age), and test performances tend to drop somewhat during holidays. Of major concern, these types of findings do not only pertain to tasks that draw heavily on knowledge, but also measures often regarded as relatively insensitive to acculturation, including tests of fluid aspects of cognition, like reasoning tests and episodic tests. Thus, in light of the fact that successive birth cohorts have been exposed to increasingly larger longer education, it is an important variable to consider when age differences in episodic as well as semantic memory are evaluated.

Typically, studies involving young-old comparisons have attempted to control for differences in educational attainment by matching groups with regard to educational background. The major problem using this method is that young and old groups that are equated with regard to educational background are likely to be differentially representative of their cohort. In addition, quantity of formal schooling (years of education) is an often used choice. It suffers from the problem that not only the quantity of education may have changed across generations, but also its quality. An alternative is to compare (categorical) levels of education (e.g., Park, 2002). This type of measure seem may control for some qualitative aspects, but ignores the potential cross-historical variation in quantity of schooling (i.e., in years) within different levels of education. Without a better means to control for educational attainment than quantity (or levels) of formal education, years of schooling is an important factor when cross-sectional age differences in memory (and cognition in general) are concerned.

Is there a better means to separate age and cohort effects? Even if education proved to be a valid indicator of cohort-related influences it is but one of an unknown number of potential cohort factors. As mentioned earlier, this may be more difficult than often conceived due to the potentially complex relation between age, time, and cohort-related influences. However, one important source of evidence is available once we can compare the age-gradients derived from cross-sectional and longitudinal studies of the same group of individuals.

**Longitudinal studies**

The most fine-grained comparisons of cross-sectional versus longitudinal patterns of changes in aspects of cognitive functioning emanate from the Seattle Longitudinal Study (SLS; Schaie, 1996). These data primarily departed from the primary mental abilities framework by Thurstone (1943/1938). Overall, these results illustrate that a very different pattern of results emerge from longitudinal as compared with cross-sectional data. Specifically, several abilities that appear to be associated with early age-related deficits from viewpoint of cross-sectional data, for example spatial ability and reasoning, instead adhere to a pattern of no or little changes up to the late 60s followed by a decline afterwards.

As was noted before, there are large-scale investigations indicating that average levels of episodic memory performance begins to deteriorate early, followed by a pattern of continuous decline. Are there reasons to believe that time-related changes will reveal a different pattern also in the case of episodic memory?
In fact several longitudinal studies appear to contradict the notion that episodic deterioration at all. Specifically, several studies found no time-related changes within groups of middle-aged or old participants in one or several episodic measures (e.g., Flicker, Ferris, & Reisberg, 1993; Hultsch, Hertzog, McDonald-Mizczak, & Dixon, 1992; McDonald-Miszczak, Hertzog, & Hultsch, 1995). In other cases, time-related improvements were observed (e.g., Bäckman et al., 1998; Flicker et al, 1993; McCarty, Siegler, & Logue, 1982; Hultsch et al., 1992, Mitrushina & Satz, 1991). As noted by Salthouse (2000), lack of power of detecting decline may be a factor in longitudinal studies. However, time-related increments in performance levels are obviously difficult to reconcile with cross-sectional findings.

Other longitudinal studies have reported time-related declines approaching those predicted on basis of cross-sectional data, though (e.g., Chistensen 2001, Johansson, Zarit, & Berg, 1992; Mayeux, Small, Tang, Tycko, & Stern, 2001; Shichita, Hatano, Ohashi, Shibata, & Matuzki, 1986). The latter set of studies often used longer intervals between measurement occasions relative than the former set of studies. A possibility is thus that the mixed evidence reflects variations in retest interval which is possibly related to susceptibility of practice effects. Another possible factor is age of the participants. A recurrent pattern in studies covering a wider age range is namely that baseline age is negatively related to memory change (e.g., Colsher & Wallace, 1991; Finkel, Pedersen, Plomin, & McClearn, 1998; Zelinski et al., 1993). Thus, a possibility is that with passage of time, young/middle-aged groups remain relatively stable whereas older groups decline.

Seven-year data from the SLS (Schaie, 1994) on “verbal memory” (a factor score primarily reflecting two episodic word recall measures) are illustrative of the foregoing pattern. In the period from young adulthood (25-28 years) to about age 60 minor improvements were observed and no detectable deficit relative to the mean for the youngest was found until age 67. Past this age, pronounced decline was observed. Similarly, Zelinski and Burnight (1997) observed no 16-year changes in list recall for a group initially 30-36 years old, whereas older groups declined at a rate similar to that predicted from cross-sectional data. If valid, these results indicate that the notions of an early onset and linear progression of decline of episodic memory functioning from early to late adulthood may not provide an accurate description of the average intra-individual changes that occur as the result of maturation.

In sum, there are several findings in the literature that challenge the view that fluid aspects of cognition are as age-sensitive as often believed. Of primary interest here, there are indications in the literature the longitudinal data on episodic memory depict a different aging pattern than cross-sectional data. To examine the viability of these trends a wider range of age cohorts need to be investigated both from the point of cross-sectional and longitudinal analyses.

Internal validity threats in longitudinal studies

Even though it is often asserted that the longitudinal design is superior to the cross-sectional with regard to the possibility of drawing firm conclusions regarding age-related
changes, several internal validity threats that are specific to longitudinal data should be paid attention to.

In a similar vein as age and cohort are confounded in cross-sectional studies, effects due to time of measurement, or events occurring between points of measurement, are confounded with effects due to aging in longitudinal studies. Most available studies sidestep this issue by assuming that cultural changes that occur during a relatively short test-retest interval have a negligible impact on performance on tests of basic cognitive processes. This may be a reasonable assumption. However, one can imagine cases in which time or period-related effects could affect such measures. Consider for example, the case that we assess the same group of individual on a test of matrix reasoning in 1998 and 2003, and that newspapers suddenly after the first measurement occasion provide their readers with matrix tests each Sunday. In this case, the conclusions regarding age-related changes in tests of matrix reasoning could be distorted by the widespread use of practice on similar problems (i.e., underestimate negative age-related changes that might be expected in such a task).

A second potential problem is that the measurement instruments, or the way they are administered, change across measurements. A third potential problem is that attrition is inevitable across a retest interval spanning several years. People drop out from longitudinal studies for a variety of reasons. However, this dropout is often not random. Specifically, returnees are often found to be better than those who dropped out in terms of baseline (initial) level of cognitive performance. This matter may reflect a variety of factors, including social factors, health-related factors, or terminal decline. The problem then, especially if there is a high frequency, is that levels of functioning observed in the returnees may not be very representative of the population intended to draw conclusions about. This problem may be especially serious if we have reasons to expect that initial level of performance is related to rate of change, e.g. such that groups of lower performing individuals (e.g., those who drop out) are expected to exhibit more decline in cognitive functioning. The relation between ability level and rate of decline is generally considered as difficult to appreciate (e.g., due to regression effects). However, on basis of cognitive studies there seems to be little reason to expect that there is an obvious relation between level of cognitive functioning and rate of change (see Kausler, 1991).

A more palpable, but almost always ignored problem in longitudinal studies is concerned with progressive errors. One source of such errors is related to the fact that the same instruments are used across measurement occasions. More specifically, taking a test once, twice or several times during the course of a longitudinal study could serve to minimize the likelihood of detecting decline in functioning, or even reverse the direction of changes from negative to positive. Most of the longitudinal studies above commented on this possibility as part of the general discussion. A recent study by Rabbitt et al. (2001) underscored the need to pay attention to practice effects in longitudinal studies of cognition by showing that practice effects on a measure of fluid intelligence may be as large as the age effects even with a retest interval of several years. Also, Schaie (1988) provided some indications that practice effects may be present across an even longer retest interval. In none of the aforementioned studies on memory control for these effects
were included.

In sum, a longitudinal design needs to be adopted to draw strong conclusions regarding the intra-individual changes that accompany aging. At the same time as internal validity threats are of serious concern in cross-sectional studies, especially cohort effects, longitudinal measurement is coupled with unique problems. Next, we shall turn to the Betula Prospective cohort study, from which the data of the present thesis emanated, focusing on major aspects of the study, including its basic design, the recruitment procedures, and the included measures of cognitive functioning.

**The Betula prospective cohort study**

The empirical studies of the present thesis were based on a longitudinal study of memory and health, the Betula prospective cohort study (Betula is Latin for birch tree; the Swedish community in which it is conducted, Umeå, is also known as the city of birch trees), or the Betula study for short. The primary objectives of the Betula study are to (a) trace the course of memory development across the adult life span, (b) detect preclinical signs and possible risk-factors of dementia, and (c) obtain pre-morbid measurement of performance in cases were brain-injuries and other damages occur. The study was originally planned to encompass three test occasions, each separated by five years (Nilsson et al., 1997).

**Design**

An overview of the design of the Betula Study is provided in Table 3, including information of the samples involved, age at date of test and birth-year.

**Table 3**

*Overview of the design of the Betula Study (first three test occasions).*

<table>
<thead>
<tr>
<th>Sample</th>
<th>ToM</th>
<th>Age at date of test</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>T1</td>
<td>35 40 45 50 55 60 65 70 75 80</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>40 45 50 55 60 65 70 75 80 85</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>45 50 55 60 65 70 75 80 85 90</td>
</tr>
<tr>
<td>S2</td>
<td>T2</td>
<td>40 45 50 55 60 65 70 75 80 85</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>45 50 55 60 65 70 75 80 85 90</td>
</tr>
<tr>
<td>S3</td>
<td>T2</td>
<td>35 40 45 50 55 60 65 70 75 80</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>40 45 50 55 60 65 70 75 80 85</td>
</tr>
<tr>
<td>S4</td>
<td>T3</td>
<td>35 40 45 50 55 60 65 70 75 80 85 90</td>
</tr>
</tbody>
</table>

**Birth-year**: -63 -59 -54 -49 -44 -39 -34 -29 -24 -19 -14 -09

value (range = 2 years)

As can be seen, three test occasions have been completed thus far (2nd column). These were conducted in 1988-1990 (First occasion; Time 1), 1993-1995 (second occasion; Time 2), and 1998-2000 (third occasion; Time 3) and were thus separated by five years. Participants in Sample 1 (S1) have been tested on all three occasions. Participants in Sample 2 (S2) were drawn from the same birth cohorts as the participants in Sample 1 (and thus matched with regard to age with S1-participants at Time 2). They entered the study at the second measurement occasion. The same is true of the participants in Sample 3 (S3) that were at the same age as the S1-participants when these were tested for the first time (i.e., 35-80 years). Both participants in S2 and S3 were assessed a second time at the third test occasion. Finally, the participants in Sample 4 were tested for the first time at the third test occasion. The original plan was to include 100 persons in each group (each cell in Table 3) the first time they were assessed. This was realizable at Time 1 (thus involving 1000 participants) and to a great extent at Time 2 (with the exception of the oldest groups in which the number of participants that were willing to participate and met the screening criteria did not fully reach the 100 mark). Furthermore, the plans regarding the size of sample 4 had to be adjusted before the test occasion so as to contain 50 individuals in each of the 12 groups (35-90). Owing to the same problem of recruiting participants in the oldest groups as in S2 and S3 at Time 2, the final number of participants in S4 was a little less than 500.

The outline of this design followed the model proposed by Schaie (1965). As we can see each row in this table allows for a standard cross-sectional age-group comparison involving 10-12 age groups. Within each of the samples 1-3, one (S2 and S3) or two vertical steps (S1) in each column, represents a longitudinal sequence.

In the terms used by Schaie, one type of sequential analysis that can be used to analyze these data involves multiple cohorts across one or several measurement occasion (e.g., S1 at T1 versus T2). This type of design is called a cross-sequential design and pits age/cohort-related changes against time-related changes. A second type of sequential design, the cohort sequential, involves two multi-cohort sequences (rows) across at least two or several measurement occasions (e.g., S1 at measurement occasion one and two, and S3 at measurement occasion 2 and 3). A third type of design, involves multi-cohort sequences (rows) from samples that differ systematically with regard to age rather than cohort, across at least two measurement occasions (e.g., S1 at measurement occasion 1 and 2 and S3 at measurement occasion 2 and 3).

Of primary concern at present, three of the empirical studies were based on cross-sectional data (rows of this table) and one included data for one of the samples (S1) for two measurement occasions (thus involving cross-sectional as well as longitudinal sequences). As was noted before, the S3 participants were matched with regard to age (and cohort) with the returnees in S1. One important difference is that the latter but not the former groups had been exposed to the same battery before at T2. Thus, the difference between these two samples established at the second test occasion should be informative of the presence of practice effects within the longitudinal sample.

The participants within each of the three samples were recruited in the same fashion
(across measurement occasion). A brief description of these procedures will be provided next. At present, the specific details concerning external attrition associated with the recruitment stage (e.g., information concerning how many needed to be contacted to obtained to the final samples, and comparisons between those who were willing to participate and those who were unwilling to participate are only) are fully analyzed only for S1. Thus, the figures for S1 are included where appropriate.

Recruitment procedure

A list of randomly sampled names in each of the cohorts to be included in the study was obtained from the population registry in Umeå, a Swedish community with about 100 000 inhabitants. Each person was first contacted by mail and informed that she/he had been randomly selected to participate in the study on a voluntary basis, and was informed about time investments and effort expected. He/she was informed that participation involved a health examination including blood-sampling, examination of memory functions and a further blood-sampling three months after the memory examination was completed, and that these examinations would be repeated at least once five years after the first examination (the participants in S4 were not informed of the latter, as only three test occasions were planned for). Anonymity and storage of results in a database were informed of in the letter.

To obtain the final Sample 1 (1000 participants) 1976 persons needed to be contacted. The 976 initially contacted persons that did not participate were grouped by Nilsson et al. (1997) into: (a) refusals (n = 481), (b) participants that could not be reached via telephone (n = 259), (c) participants reporting themselves to be too ill to participate (n = 130). Seventeen persons were replaced because they completed the health-exam only and did not show up on the second test session. In addition a fourth category of individuals had to be replaced due to the fact that they did not meet the specified entry criteria. This category included persons suffering from mental retardation or suspected dementia (n = 26), persons with a sensory handicap (n = 28), another language than Swedish as their native tongue (n = 35), and persons who failed to show up on the second test session (n = 17).

Comparisons of data on participants and refusals at T1 were made on demographic parameters, including gender, marital status, educational level, employment, profession, category of profession, number of children living at home, type of housing, and overcrowding. No differences between participants and non-participants in the background variables and between participants and values for the Swedish population as a whole were found, except that participants were employed to a greater extent and had a higher income than non-participants, an effect restricted to three four and three age cohorts, only. As concerns the comparison of participants and the total Swedish population, no differences were detected except that participants had a slightly higher income overall and were somewhat more educated than the Swedish population as a whole. This matter likely reflects the fact that Umeå was established as a university-city in the 1960s.
Another important aspect of attrition is that occurs between measurement occasions. Such attrition analyses are presented in detail in Study IV. In general the rates of attrition were low as compared with prior large-scale longitudinal studies of cognition (< 13%) with evidence that the attrition was selective with regard to age and cognitive measures, but not with regard to sex and education. Next, a summary of participant characteristics will be provided for S1 at T1 (see Study II and IV for background characteristics of the other samples involved in the empirical studies).

**Participants**

Individuals with a severe visual or auditory handicap, mentally retarded or demented persons and persons with another native tongue than Swedish were excluded from the sample. A majority (>70%) of the participants in each cohort reported that they felt healthy as determined by a simple yes/no assessment, one exception being the 55-year old cohort in which 60% reported that they felt healthy). Objective indicators of health confirmed these reports (Nilsson et al., 1997). Most participants in each of the cohorts were living in their own home (2% of the 80-years old lived in a nursing home). Table 4 depicts average values for the included age groups on a select number of background variables, including sex distribution, education, vocabulary, a measure of global intellectual functioning, often used as a screening test for dementia: Mini-mental State Examination (MMSE; Folstein, Folstein & McHugh, 1975) and Block-Design (Wechsler, 1981), which is regarded to reflect fluid (visuospatial) ability.

<table>
<thead>
<tr>
<th>Age</th>
<th>Females/Males</th>
<th>Education</th>
<th>Vocabulary (Max=30)</th>
<th>MMSE (Max =30)</th>
<th>Block Design (Max = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>50/50</td>
<td>13.9 (2.6)</td>
<td>23.1 (4.7)</td>
<td>28.6 (1.2)</td>
<td>34.9 (9.4)</td>
</tr>
<tr>
<td>40</td>
<td>52/48</td>
<td>13.5 (3.5)</td>
<td>23.0 (4.1)</td>
<td>28.5 (1.4)</td>
<td>31.8 (8.2)</td>
</tr>
<tr>
<td>45</td>
<td>54/46</td>
<td>12.7 (4.2)</td>
<td>22.9 (4.6)</td>
<td>28.2 (1.6)</td>
<td>32.1 (9.1)</td>
</tr>
<tr>
<td>50</td>
<td>60/40</td>
<td>10.4 (3.7)</td>
<td>22.9 (4.0)</td>
<td>28.2 (1.6)</td>
<td>29.9 (8.5)</td>
</tr>
<tr>
<td>55</td>
<td>54/46</td>
<td>8.9 (3.3)</td>
<td>22.1 (5.0)</td>
<td>28.3 (1.6)</td>
<td>29.6 (5.4)</td>
</tr>
<tr>
<td>60</td>
<td>46/54</td>
<td>8.8 (3.2)</td>
<td>21.4 (4.8)</td>
<td>28.1 (1.6)</td>
<td>25.7 (7.8)</td>
</tr>
<tr>
<td>65</td>
<td>54/46</td>
<td>8.2 (2.9)</td>
<td>20.3 (5.8)</td>
<td>27.7 (1.6)</td>
<td>22.8 (8.9)</td>
</tr>
<tr>
<td>70</td>
<td>48/52</td>
<td>8.2 (3.3)</td>
<td>20.1 (5.3)</td>
<td>27.6 (1.6)</td>
<td>22.5 (8.8)</td>
</tr>
<tr>
<td>75</td>
<td>53/47</td>
<td>7.5 (2.8)</td>
<td>17.9 (6.4)</td>
<td>26.8 (2.5)</td>
<td>17.2 (8.2)</td>
</tr>
<tr>
<td>80</td>
<td>59/41</td>
<td>7.4 (3.1)</td>
<td>17.7 (6.4)</td>
<td>26.2 (2.4)</td>
<td>14.0 (7.6)</td>
</tr>
</tbody>
</table>

Note. SDs within parentheses.

As can be seen in Table 4, the number of women exceeded that of men with a total of 540 women (and 460 men) across age groups. This reflected the distribution of target population. Quite substantial age-related differences with regard to quantity of educational attainment are evident across age groups, favouring the young (or later-born)
participants. It might be noted that these differences are largely attributable to differences among the five youngest groups. An MMSE score of 23 (or 24) or below is often considered a possible sign of dysfunction (e.g. dementia). As we can see, the averages are all well above this criterion. As regards the Block Design test it is interesting to note that these data are very similar to the age trend for raw scores evident from WAIS-R norms. The mean scores in Table 4 are about two points higher across the age cohorts than would be expected from these norms. This small difference (roughly 0.3 SD) may be related to subtle differences in sampling procedure between studies, differences between the target populations or some other factor (e.g., time lag) but the close adherence in mean levels is nevertheless noteworthy.

Data collection procedure and memory measures

The participants were examined on two separate sessions, one primarily devoted to health-examination (first session) the other to testing of memory functions (second session). The exams were taken in that order, and were separated by a week. In S1 at T1, most of the participants ($n = 912$) took part in the health and the memory examination in laboratories at the Department of Psychology, Umeå University. The remaining participants could not come to this exam, and were therefore tested in their homes and, in a few cases, at institutions for the elderly. An examination of the corresponding frequencies across the other samples and measurement occasions reveal similar proportions.

At the first session the participant met a nurse, and underwent a health examination. They were also interviewed about their health status, and received a questionnaire about social and economical issues as well as a questionnaire of critical life events. In addition a few cognitive tests were given. The next exam was undertaken by a trained experimenter who administered a battery of cognitive tests, primarily concerned with memory functions. The cognitive measures included at each of the two exams are listed in the order they were administered in Table 5. The reader interested in further details concerning the individual cognitive measures may consult the individual studies and Nilsson et al. (1997).

The list of tests in Table 5 holds for the first measurement occasion (T1). At the second measurement occasion some minor modifications of the battery were undertaken. Specifically, a word-fragment test based on words from the vocabulary test was administered as part of the first test session. In addition a planning test in form of the Tower of Hanoi was included at the end of the second test session (no. 22 in the list above, i.e. just before the activity memory test).

At the third measurement occasion (T3) a letter-digit task assumed to reflect perceptual speed was included and was administered before the Tower of Hanoi task. In the recognition test of nouns from enacted/nonenacted sentences (no. 15) the old/new responses were followed by instructions related to the Remember/Know judgment procedure that, as described previously, required the participant to provide a judgment for each item he/she had reported as “old”. Finally, one measure, the item/source recall test previously administered by the nurses was omitted from the battery.
Table 5
Cognitive tasks administered at the first and second sessions (at Time 1) in order of presentation.

<table>
<thead>
<tr>
<th>Order</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>First session</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Vocabulary test</td>
</tr>
<tr>
<td>2.</td>
<td>Study episode for recall of statements/source</td>
</tr>
<tr>
<td>4.</td>
<td>Mini Mental-State Examination (MMSE)</td>
</tr>
<tr>
<td>5.</td>
<td>Draw-a-man test</td>
</tr>
<tr>
<td>Second session</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Instructions for a later test of prospective memory (22)</td>
</tr>
<tr>
<td>2.</td>
<td>Study episode for a later face recognition test</td>
</tr>
<tr>
<td>3.</td>
<td>Study episode for a later name recognition test</td>
</tr>
<tr>
<td>4.</td>
<td>Study and immediate free recall of (enacted/nonenacted) sentences, first list</td>
</tr>
<tr>
<td>5.</td>
<td>Study immediate free recall of (enacted/nonenacted) sentences, second list</td>
</tr>
<tr>
<td>6.</td>
<td>Cued recall of nouns in sentences presented in 4 and 5</td>
</tr>
<tr>
<td>7.</td>
<td>Stem completion of surnames presented in 3</td>
</tr>
<tr>
<td>8.</td>
<td>Word fluency: initial letter A</td>
</tr>
<tr>
<td>9.</td>
<td>Word fluency: initial letter M for five-letter words</td>
</tr>
<tr>
<td>10.</td>
<td>Word fluency: initial letter B for professions</td>
</tr>
<tr>
<td>11.</td>
<td>Word fluency: initial letter S for five letter names of animals</td>
</tr>
<tr>
<td>12.</td>
<td>Yes-no recognition of faces presented in 2</td>
</tr>
<tr>
<td>13.</td>
<td>Multiple-choice recognition of names presented in 3</td>
</tr>
<tr>
<td>14.</td>
<td>Block Design test</td>
</tr>
<tr>
<td>15.</td>
<td>Yes-no recognition of nouns in sentences presented in 3</td>
</tr>
<tr>
<td>16.</td>
<td>Cued recall of nouns in sentences presented in 4 and 6</td>
</tr>
<tr>
<td>17.</td>
<td>Source recall of sentences presented in 4 and 6</td>
</tr>
<tr>
<td>18-21</td>
<td>Study/recall of word-list with or without concurrent task (four conditions)</td>
</tr>
<tr>
<td>22.</td>
<td>Incidental recall of activities in whole session (1-21)</td>
</tr>
<tr>
<td>23.</td>
<td>Prospective memory test</td>
</tr>
</tbody>
</table>

Some prior findings

To date a considerable number of studies have been published based on data from the Betula study, most of which were based on cross-sectional data. At present a few word will be said concerning the relations between age, memory, and indicators of health, an issue that was not examined within the scope of the empirical studies of this thesis.
As part of the analyses of data on episodic and semantic memory and on a measure of priming, Nilsson et al. (1997) examined the relation between memory and 16 health indicators using data the S1-sample at T1. One of the health indicators was a dichotomous subjective health rating, 10 of the indicators were measures derived from blood-samples (e.g., cholesterol level, triglyceride levels). In addition, the relation between memory performance and measures of systolic and diastolic blood pressure, number of visits to doctor, and number of medications taken were included. Finally, a vision-reading measure was included. As judged from zero-order correlations based on the full sample there was a significant ($p < .01$) association between memory and a majority (12) out of the 16 health indicators for the episodic measure, but in not for measures of semantic memory and priming. All of the significant association were low and indicative of a negative relation between health status and memory (ranging from $r = -.15$ to $r = -.34$). Of primary concern none of the corresponding associations attained a significant level when age was controlled for (partial correlations ranging from -.09 to .06).

On the basis of these results (and hierarchical regression analyses showing that age still accounted for substantial amounts of variance when all of the health indicators were entered before age), the conclusion drawn was that the health-memory relation was completely mediated by age, and thus that only a little of the relation between age and memory reflect health-related factors. The same lack of relation between age and health factors has been reported in other publications, using individual memory measures in Table 5 as the dependent measure. This included several of the retrospective episodic measures (e.g., Erngrund, Mäntylä, & Nilsson, 1996), as well as the measure of prospective memory (Mäntylä & Nilsson, 1997). This matter has been attributed to the fact that most of the observations on the health indicators were in the normal range. That is, despite indications that some of these health-markers have been found to have an influence in several other studies (for a review, see Bäckman et al., 2000) the impact of these factors may be relatively small, provided that most of the participants are in good health, something that is likely associated with the adoption of the aforementioned screening criteria.

Of further interest, previous investigations have found a gender effect for the episodic measures, showing slightly higher levels of performance in women as compared with men (Herlitz et a., 1997), a finding in line with several past studies. Importantly, there was, in line with the aforementioned study by Salthouse and Meinz (1995) no age by gender interaction, suggesting that the age trajectories for men and women do not differ.

**Research objectives**

The overarching objective of the thesis was to examine declarative (episodic, semantic) memory functioning across the adult life span. The individual studies encompassed three major issues. The first concerned the possibility that encoding support in the form of overt motor activity, enactment, modifies the magnitude of age-related differences in episodic memory. The evaluation of this issue included analyses of measures of item memory (free
recall, cued recall and overall recognition) as well as phenomenological aspects associated with memory retrieval (remembering and knowing). As was concluded from the prior review of studies, the initial results favoring the possibility that enactment may serve to compensate for age-related deficits in recall is not supported by the majority of subsequent data. Less evidence is available concerning recognition and the fact that studies using the remember/know paradigm are generally more supportive of a production deficiency account. The second issue, addressed by Study III and Study IV, concerned the organization of declarative (episodic, semantic) memory from the point of a wider range of measures. Of interest was to provide a further test of the viability of the distinction between episodic and semantic at the ability factor level (Nyberg, 1994), to investigate the possibility that these factors are decomposable into sub-factors (Study III), and to examine whether the established model(s) were age invariant. Of further interest was to appreciate the age-related variations that occur among the putative ability factors. The third major issue, addressed by Study IV concerned whether the age-performance trajectories established on basis of cross-sectional and longitudinal data at the level of second-order ability (episodic and semantic) factors would differ.

**SUMMARY OF THE EMPIRICAL STUDIES**

A summary of each of the four empirical studies of the present thesis is provided below. The first two studies focused on the relation between encoding support in form of enactment and age differences in memory, as assessed by various tests of item memory and subjective reports of recollective experience. Study III was devoted to the ability structure of declarative memory and the patterns of age-related differences of the factors identified. In Study IV cross-sectional and longitudinal mean-level aging patterns were compared using composite measures of episodic and semantic memory.

**Study I**


Study I examined age differences in recall of lists of sentences encoded with or without enactment. The data emanated from Sample 1 at first time of measurement and included 1000 participants (35 to 80 years old). The study phase involved two separate lists of 16 verb-noun sentences, one for each encoding type (heard/read only or enacted). Each noun within the lists belonged to one of four semantic categories. Memory of the sentences/nouns was assessed by two tests: immediate free recall and category-cued recall. In the first test, which followed the last item of each of the two lists, participants were requested to orally recall as many of the sentences as possible. In the latter test participants were provided with the superordinate categories into which previously studied items belonged to be used as cues for retrieval. In addition degree of clustering
was measured to examine the extent to which subjects utilized the semantic list structure in free recall. As a complementary measure of events involving performance on part of the participants, age differences in incidental recall of a series of cognitive test activities engaged in during a longer test-session (i.e., the tests listed in Table 5, second exam) was examined. The results revealed a substantial enactment effect, that is superior recall of enacted as compared with nonenacted sentences across the tests. However, age differences were similar in magnitude for enacted and nonenacted sentences. The weak interaction in analyses of the overall free-recall data that was in the opposite directions to what would be expected from a production deficiency/compensation hypothesis (Bäckman & Nilsson, 1984) possibly reflected a scaling (floor) effect for nonenacted sentences as was indicated by median-split analyses. No tendency of an age by encoding interaction was observed in cued recall. Thus, enactment boosted recall substantially, but this form of encoding support appears to be utilized to about the same extent regardless of age. Clustering was marginally higher for recall of nouns from the enacted condition but was unrelated to age, suggesting that age-related differences observed in recall reflected age-related problems of encoding/retrieval of item-specific information rather than of relational information. Negative age differences indicative of a gradual deterioration of performance with advancing age were further observed in activity recall. The magnitude of age differences was as large for cognitive tasks involving motor activity as for tasks not involving motor activity, in line with the results for enacted/nonenacted sentences. Analyses of the potential effect of differences in educational attainment, finally, suggested that part of the age-related trends across the memory measures may reflect cohort differences. Specifically, there was no evidence of unique age-related variance in memory across the five youngest cohorts when education was entered before age in a hierarchic regression analysis. By contrast, significant amounts of age-related variance remained in the corresponding analysis involving the five oldest cohorts, suggesting that genuine age effects in episodic memory may be more prominent in old age. It was concluded that changes in some more fundamental factor(s) than the nature of the encoding or retrieval tasks must be sought for to explain the deficits in episodic memory measures that occur in old age.

**Study II**


The objective of Study II was to investigate age differences in recollective experience accompanying recognition from point of the Remember/Know-paradigm. The cross-sectional data were drawn from Sample 4 at third time of measurement and the study involved 321 participants ranging in age from 35 to 95 years (the R/K procedure was not introduced at the start of the third data collection wave, hence the smaller $n$). Recognition of nouns studied under enacted/nonenacted encoding conditions were compared to test the
hypothesis that diminished level of remember experiences observed in groups of older adults is compensated for by provision of support or guidance (Perfect et al., 1995). Level of overall recognition performance was in addition examined, using A’ as the dependent measure, in addition to hits and false alarms. The R-judgments were subjected to a corresponding series of analyses. The relation between age differences in remembering and overall recognition and a composite measure of processing speed was furthermore examined. The latter was based on two measures: number of completed items in the letter-digit substitution task (1 min.), a modified version of the digit-symbol measure of WAIS-R, and time to solve a very easy item from the Block Design test (Wechsler, 1981). The results demonstrated higher levels of remembering following enacted encoding in line with past studies (Engelkamp & Dehn, 1997). Age differences consistent with a gradual age-related deficit were observed in overall recognition (A’) and frequency/discrimination (A’) of R judgments across conditions. As judged from the overall data, these differences were somewhat larger in magnitude in the nonenacted condition. However, the performance levels and levels of R-judgments approached ceiling level across the youngest groups in the enacted condition, challenging the interpretation of the age by encoding interaction. In fact, in attempts to remedy a potential ceiling effect by considering below median-split data (based on scores in the enacted condition), the interaction fell short of significance suggesting that the interaction effect observed in the analysis of the overall data may have reflected a scaling (ceiling) effect. Know levels were overall low in the enacted condition (cf. Engelkamp & Dehn, 1997) and the pattern of age-related increment of Know judgments was difficult to interpret given the high hit-rate for this condition. The frequency of know-reports exhibited an age-related decrease in the nonenacted condition but this effect was weak. Thus, the results of the present study failed to support the production deficiency hypothesis proposed by Perfect et al. (1995) in two ways: (a) in that the magnitude was largely invariant across encoding type, and (b) in that the aging pattern for K-judgments, if anything, exhibited a tendency of a pattern reversed to that expected from this hypothesis.

By means of a portioning of variance-technique (commonality analyses), the relation between recollective experience (and overall recognition) and processing speed was investigated. Years of educational attainment was also entered. Results showed that large amount of the age-related variance was shared by processing speed, both as regards levels of R-judgments and as regards overall recognition. This was taken to suggest that the age-related variance in remembering and overall recognition reflects age-related reductions in processing speed or, alternatively, reductions in a factor common to speed and memory. Taken together, these results were consistent with a common cause or processing efficiency interpretation, rather than with a production deficiency interpretation of age differences in recollective experience in showing that the age-related variance in an internal factor (speed), rather than presence or absence of specific encoding instructions control age differences in frequency of recollective experience and overall recognition.
Study III


Study III was based on cross-sectional data for a total of 925 participants from Sample 3. Competing models of declarative memory were compared from the point of covariance structures among a variety of measures in the battery classifiable as episodic and semantic. A middle-aged group (35-50 years; \( n = 377 \)) was used to establish a best fitting model. Next, this model was tested for age-invariance, e.g. the extent to which the identified factor structure generalized across groups of young-old (55-65 years; \( n = 283 \)) and old-old (70-80 years; \( n = 265 \)) participants. The results of the initial analyses confirmed the results of a previous study by Nyberg (1994) in that a two-factor (episodic-semantic) model yielded a better fit than a unitary, or one-factor, model. Based on findings in the literature indicative of a further differentiation of episodic and semantic memory into two sub-factors: recall versus recognition (episodic memory) and knowledge versus fluency (semantic memory), a subsequent comparison was made between the aforementioned two-factor model, and models involving the other factors (one in which the two potential sub-factors were regarded as second order factors subsumed by the episodic-semantic factors and one without first-order factors). The fit indices indicated that these models represented an improvement over the two-factor model. The six-factor model with episodic and semantic memory as the second-order factors and recall versus recognition and knowledge versus fluency as the first-order factors was judged to be more tenable from a theoretical point of view. In a next step, this model was tested for age invariance. From point of several indices, this model was found to be metrically age-invariant. Specifically, the factor loadings did not differ among groups (metrical invariance). This suggests that meaningful age-comparisons may be made along the factors identified.

With regard to the factor-specific age-groups differences, the results were consistent with several prior studies in showing relatively small age-related variations in semantic memory performance together with pronounced negative age-group differences in episodic memory. Specifically, the middle-aged group outperformed the young-old group that in turn performed at a higher level than the oldest group on the episodic factor. As concerns semantic memory, the data instead indicated that semantic memory improves into young-old age, with evidence of some decrement in old age. It should be noted that the magnitude of this apparent deficit was of about the same magnitude as the apparent age-related improvement up to young-old age. Thus, the youngest and oldest groups did not differ on the semantic factor.

Of further interest were the patterns of age-related differences discernible on the first-order factors. At this point, the results revealed some evidence of selective age
differences within the episodic, and to some extent, the semantic domain. As concerns the episodic factor, a pattern of lower performance in the young-old and old-old groups as compared with the middle-aged group was observed in recall, which confirms a variety of other studies (e.g., Craik & McDowd, 1987). By contrast, on the recognition factor only the old-old group was found to perform at a lower level as in comparison with the middle-aged group. As concerns the semantic sub-factors, a positive age difference in knowledge between the middle-aged and young-old group was observed, together with a minor but significant age-related deficit in the oldest group (in comparison with the young-old group). By contrast, fluency did not differ between the middle-aged and young-old group. There was a tendency for fluency to be lower in the oldest group as compared with the young-old group. Taken together, these findings confirm, and extend, previous findings that episodic memory is associated with larger age-related variations than semantic memory, and that recall may be more susceptible to negative age effects as compared with recognition. The variability within the semantic domain is suggestive of a late peak-level age for knowledge, and some impairment in old-old age, and minor age-related changes in fluency.

**Study IV**


In study IV five-year changes in composite measures of episodic and semantic memory performance, based on ten different tests (five for each of the two memory factors), were examined in a sample of 829 participants from 10 age cohorts (35-80 years). A second, cohort-matched, sample; S2 (n = 959) was assessed at Time 2 to control for practice effects. Changes observed across the five-year interval were indicative of improvements in episodic memory performance for the younger cohorts, in contrast with the gradual age-related decrements depicted by cross-sectional data for S1 and S2. However, within age groups 60 years or older initially average performance levels declined substantially, at a rate that approximated that expected from the cross-sectional data. Minor practice effects were observed for the episodic measure, but even when these were adjusted for results were consistent with a late onset (> 60 years) of decline for episodic memory. Specifically, a plateau in performance was discernible from the composite age-gradients, followed by substantial decline in old age. Semantic memory showed minor improvements until age 55, with less steep decrements as compared with episodic memory in old age as judged from the composite longitudinal age-gradients. Thus, the basic difference with regard to the magnitude of memory impairments for semantic and episodic memory observed in Study III was replicated. To examine whether differences between the longitudinal and cross-sectional age-gradients could reflect cohort differences in educational attainment hierarchical regression analyses were conducted on the cross-sectional data. The results of these analyses were taken to suggest the possibility
that part, perhaps most, of the difference between cross-sectional and longitudinal aging patterns is accounted for by variations across cohorts with regard to quantity of formal schooling. Specifically, in the case of episodic memory, no unique age-related variance remained across the five youngest cohorts when age differences when education was entered before age, in line with the longitudinal data (cf. also Study I). For semantic memory, the corresponding analyses were suggestive of minor improvements in performance in line with the pattern depicted by the longitudinal data, once education was partialed out. Both in case of the episodic and semantic memory measures, unique age-related variance remained for the older cohorts, more so for the episodic measure, also in line with the longitudinal patterns. Collectively, these observations were taken to suggest that the course of episodic and semantic memory development in adulthood differ, and underscore the need to control for internal validity threats in cross-sectional as well as longitudinal studies. The results of maintenance of declarative memory functioning into late middle-age, at about age 60, with decline thereafter, was discussed in light of corresponding patterns of changes in potential mediator variables.

**GENERAL DISCUSSION**

The overarching objective of the present thesis was to examine episodic and semantic memory functioning across the adult life span. In the following section major results from the four empirical studies will be discussed. First, the issue of utilization of support in episodic memory will be discussed from viewpoint of the results in Study I and Study II. Thereafter, the aging patterns on episodic and semantic memory and putative sub-factors will be commented in some more detail, followed by a discussion of cross-sectional versus longitudinal aging patterns and their implications for theoretical accounts of cognitive aging. Finally, some limitations of the results will be mentioned, some future research steps will be outlined and, a few concluding remarks will be made.

**Utilization of cognitive support**

To recapitulate, variants of a production deficiency hypothesis posit that provision of cognitive support should reduce the negative age-related differences in episodic measures observed under non-guided, or unsupported, conditions. Encoding enactment, the specific form of encoding support examined in the present thesis has been suggested to serve as such a compensatory device (e.g., Bäckman & Nilsson, 1984, 1985). However, the present results (Study I and II) extend the more generally obtained finding across previous studies review earlier by showing parallel, or largely parallel, age-related differences in memory of materials encoded with and without such support. The generality of the foregoing findings appear to be fairly high provided: (a) that this finding was obtained both in free recall, category-cued recall (Study I), recognition (Study II) and verb-cued recall (Erngrund, Mäntylä, & Rönnlund, 1996), (b) the fact that these patterns were observed across a wide age range, also including participants older than those examined in past studies (35-80 in Study I, 35-90 in Study II), and (c) indications that cognitive activities showed a
similar age-related pattern in recall regardless of involvement of motor activity (Study I).

The data in Study II were used to test a more specific version of a production deficiency hypothesis, namely that age-related differences in the frequency of recollective experiences (R-judgments) accompanying recognition may be minimized provided that encoding is guided by specific instructions (Perfect et al., 1995). Specifically, the lower levels of remember experiences often observed in groups of older adults have been hypothesized to be selectively enhanced in older adults provided that encoding is guided (e.g., in form of specific instructions). As indicated by a review of past studies the published data seem to be more favorable to this specific production deficiency hypothesis than that concerned with memory for item recall or recognition.

However, the result of Study II failed to support this hypothesis by showing lowered levels of remember judgments with advancing age irrespective of presence or absence of enacted encoding support. Apart from advantages in comparison with past studies in terms of age range, degree of representativeness, and size of the included samples, it should be re-emphasized that the pattern in some prior studies that offered some support of a production deficiency interpretation, emerged from considering differences across, rather than within, studies. Thus, the present observations provide fairly strong evidence of a pattern of generalized negative age-related effects within the domain of episodic remembering in old age. Future research will be required to resolve the discrepancy between the extant studies with regard to the effects of encoding support on age differences in remembering. Scaling factors, such as near ceiling level performance in the guided conditions for the younger adults could be a factor in addition to procedural variations across studies. A possibility, of course, is that forms of support other than enacted encoding have the potentials to reduce age-related differences.

As plausible reason for the persistence of age-related differences in memory of enacted information is that the sensory-motor information that is encoded by overt activity increases the distinctiveness of the encoded events without altering the demands on age sensitive cognitive/neural resources critical for successful encoding. The finding that a very high proportion of the age-related variance in recollective experience and overall recognition was shared with/predictable from individual differences in cognitive speed (Study II), is interesting in this regard, regardless of whether it signals a direct influence of speed on age differences in recollective experience or whether this result is due to the fact that both measures are dependent on some common age sensitive factor. In any case, the high degree of association between speed measures and age effects in recollective experience agrees with the results of another recent study (Clarys, Isingrini, & Gana, 2002). In addition to speed, the foregoing study examined the influence of working memory on age differences in recollective experience. A structural equation model in which all of the age-related influence on frequency of recollective experience was mediated by individual differences in speed, via reductions in working memory, provided a good fit. However, as judged from other sources of data (e.g., Park et al., 1996; 2002), there is good reason to believe that working memory capacity predicts age-related variance in episodic memory over and above speed. Also, the results by Earles (1996) are highly
relevance at present, as measures of working memory were at least as predictive of the age-related variance in recall of enacted/nonenacted materials as were measures of perceptual speed. As was noted initially, the relation among markers of cognitive measures as regards the role of mediators is currently uncertain. To the extent that Block Design is representative of a general cognitive factor it might be noted that individual differences in this measure was a strong predictor of the age-related variance in recall of enacted and nonenacted tasks (Erngrund et al., 1996; see also Nyberg et al., 1996).

Even though the impact of basic processing factors and their neurological correlates may be less likely to be remedied by interventions, the present lack of evidence that cognitive support in form of enactment reduces negative age differences in memory should be judged in light of the following observations: (a) that the levels of memory performance was enhanced substantially even in the oldest groups, that is, the elderly show a well preserved plasticity, or ability to utilize the support, and (b) that it is the individuals that tend to lack adequate resources to perform well under non-guided conditions, such the elderly, that are most in need of support. Viewed from this perspective, it is interesting to note that groups as old 90 years approached the levels of non-guided younger adults given provision of enacted encoding support (Study I and Study II). To the extent that cognitive support in form enactment may serve as a basis for interventional strategies this may be the more important aspect of the results than the issue of whether or not support reduces the age differences in memory per se. From the view that enactment increases item distinctiveness a hypothesis that remains to be tested is whether it reduces associatively based false memories. As older adults tend to show higher rates of such false remembering (e.g., Koutstaal & Schacter, 1997) such a result could, in an analogous manner, be important whether or not the predicted reduction in the rates of false memories of the older adults reach the same levels as that of younger adults.

Finally, with regard to the effect of cognitive support, at least one aspect of the results, namely the finding of less pronounced age-related differences in recognition as compared with recall (Study III), may be taken to suggest that support reduces the magnitude of age differences in episodic memory (cf., Spencer & Raz, 1995). At the same time, substantial age-related differences were obtained also when retrieval cues were provided at times of retrieval and when recognition served as the dependent measure (Study I, II, III, and IV). Thus, in line with the dominant picture in the literature (Bäckman et al., 2001), the present findings indicate a weak relation between age differences in episodic memory and degree of cognitive support.

**Factor-specific aging patterns**

The present results converged on a pattern of substantial losses in memory functioning in old age (Studies 1-IV). However, one important aspect of the results of Study III and Study IV is that performances on a wider set of declarative memory measures show factor specific age gradients. In line with the bulk of previous research episodic and semantic memory measures were associated with quite different age trends, both as judged from the cross-sectional age differences and longitudinal data, consistent with less pronounced decline in old age in the case of semantic memory. Whereas this
was a predictable finding, much prior evidence regarding the issues of differential aging patterns in episodic versus semantic memory draw on patterns evident across rather than within studies. More importantly the present patterns were derived from a factor-analytic approach, which should yield a better means to appreciate the latent constructs and the models were tested for age-invariance before examining the factor-specific aging patterns which is still rare in cognitive aging research.

More novel findings with regard to the organization of memory abilities (Study III) were (a) the finding of sub-factors within the episodic as well as the semantic tasks domain, and (b) the finding that there is some variability with regard to aging patterns among these sub-factors. Starting with episodic memory, the findings were consistent with larger age differences in recall as compared with recognition, a finding in well agreement with past research (Craik & McDowd, 1987; Spencer and Raz, 1995), but which, to my knowledge, never has been demonstrated before at the factorial (latent mean) level. As discussed before, recall and recognition measures differ in the extent of retrieval support offered; more in the case of recognition. Thus, in recognition the demands on attentional resources/working memory capacity that are likely to be diminished in old age are may be smaller. To this point, there are findings consistent with the idea that working memory capacity mediates the age-related variance in episodic memory over and above influences of speed in recall, but not in a less demanding episodic measures, including recognition (Park et al., 1996; Whiting & Smith, 1997). In a related vein, results conducted from secondary task paradigms show that the attentional demands are larger in test of recall than recognition, in particular for older adults (e.g., Craik & MacDowd, 1987). Another explanation of the differential age sensitivity of recall and recognition measures is that they rely on recollection and familiarity to a different extent (Jacoby, 1991), and that older adults are able to manage fairly well on recognition measure due to spared familiarity-based retrieval processes. Future research will be needed to investigate the underpinnings of the age-related effects in recall versus recognition further but the present observations are consistent with the view that recognition measures are less age sensitive than recall measures, presumably due to differential processing demands.

Turning to semantic memory, the cross-sectional (education adjusted) age differences and the aggregated patterns of time-related changes converged on a pattern of (a) small but reliable improvements from age 35 to about age 60 (Study III, Study IV), and (b) reliable deficits in old age (Study III and especially Study IV). Whereas the estimated changes in semantic memory across the adult life span were small in comparison with those observed for episodic memory, especially with regard to the magnitude of decline in old age, some age-related variation is observed and need to be accounted for.

Any model of age-related changes in semantic memory must probably take into account that advancing age is associated with increased opportunities to acquire factual and verbal knowledge (cf., Salthouse, 1991). Perhaps this accumulation of semantic knowledge may be expected to decelerate somewhat in old age due to age-related deficits of encoding semantic information (Service & Craik, 1993) and, perhaps, lower levels of engagement in cultural activities in old age (Anstey & Christensen., 2000). On the other hand, retrieval of information stored in semantic memory as assessed by measures of verbal meaning and, perhaps more so, measures
such as fluency, likely draw on basic processing or neural capacities. Specifically, these tasks place demands on rapid responding, inhibition of no-longer relevant responses, reasoning et cetera. Thus, at a certain point in life, about age 60 as judged from the present results, the positive contribution associated with increased age in form of added cultural experience may be outweighed by the negative influence of the “mechanics” or fluid aspects of cognition (cf., Baltes, 1987; McArdle et al., 2000). A more specific model that may be worth testing further, perhaps using an even broader set of memory measures, is thus that age-related differences or changes in semantic tasks is jointly determined by knowledge, partly by processing efficiency (e.g. as indexed by speed, working memory, executive functions) and that the relative dependency on processing efficiency in production or fluency tasks will be associated with larger age-related deficits.

To my knowledge only a few attempts to test versions of a dual process (processing efficiency versus knowledge) model in a more systematic fashion exist in the literature. The results by Salthouse (1993b) did not offer consistent support for the idea that knowledge compensates for age-related deficiencies in speed on verbal ability measures (cf. also Bryan, Luszczs, & Crawford, 1997). On the other hand, findings that the relation between age and word fluency may revert to a positive one with, but not without, statistical control of speed (Schaie, 1989) is in line with the predictions of such a model. Also the patterns in study III and Study IV, showing some age-related increments in the knowledge-factor (cf. also Verhaeghen, 2003) together with lack of such increment across the period from middle aged to young old age on the fluency factor, appear to be in line with at least some of the expectations from the outlined dual process account. Specifically, the drop in performance from the young-old group to the oldest group was larger (and only significant) for knowledge. At the same time, the present data show a less marked directional divergence with regard to age gradients in knowledge (comprehension) and fluency measures as compared with some prior studies (e.g., Schaie, 1996). One possible caveat in some of the earlier studies that applies also to the present ones is that it is difficult to assess knowledge without requiring some demands on processing resources or fluid reasoning.

Finally, a few words may be said on the differences and similarities between the aging patterns in episodic and semantic memory. Based on evidence that these two forms of memory draw on at least partly different neural resources it appears that the age differential patterns are accounted for at least partly by the fact that these neural underpinnings are differentially affected by the aging process. Also, the two forms of memory might be expected to be differentially sensitive to acculturation (more so in the case of semantic memory; cf. Anstey & Christensen, 2000, see also Study IV). On the other hand similarities between them exist and it is important to note that (a) both forms of memory show some deterioration in very old age, and, (b) that the onset of decline occurs at about the same age for both forms of memory, that is about age 60. A likely reason for this matter is that episodic and semantic tasks draw one a limited set of central age sensitive processing components (e.g., speed, executive functions and their neural underpinnings). Along the line of reasoning that unique as well as common age sensitive components of processing are required for successful performance in the two forms of memory tasks, one should expect that the longitudinal covariance patterns reveal a substantial but not perfect correspondence between
changes in the two memory factors. Another valuable source of evidence pertaining to
the hypothesis of system specific and common age sensitive components among
measures of episodic memory and measures of semantic memory could be derived
from neuroimaging data (cf., Nyberg et al., 2003) comparing groups of middle-aged
and very old participants.

Cross-sectional versus longitudinal aging patterns: Some reflections

As concerns episodic memory, the outcome of the cross-sectional results pertaining to
enacted/nonenacted tasks and activity memory (Study I and Study II) together with
those of the other individual measures (Study III, IV) were suggestive of a linear
progression of memory decline starting from age 35 throughout the remainder of the
lifespan. As noted before, this is in agreement with the bulk of past cross-sectional
studies (e.g., Park et al., 2002) and most textbook descriptions of the age-performance
relationship in episodic memory measures. One important result of the empirical
studies of the present thesis, Study IV in particular, is that such an “over the hill
model” with a very early onset of decline may not constitute a good description of the
mean level changes that occur within groups of individuals. Instead, the age-gradients
established by the present longitudinal data was consistent with a plateau in level of
performance in the period from young adulthood (age 35) to around age 60. This was
so even when retest effects were taken into account. Past this substantial time-related
losses in performance were observed.

As noted before, some prior longitudinal studies anticipated the pattern of a fairly
late onset of age-related changes in episodic memory (e.g., Schaie, 1994). As such the
patterns of Study IV also converge with findings of a late onset for other fluid abilities
regarded to decline early on basis of cross-sectional age differences, like reasoning,
and spatial orientation (Schaie, 1996). Preliminary analysis of the Block Design
measure, which, as noted before, is regarded as a highly “g”-saturated measure, appear
to exhibit a similar pattern, with stable levels of performance for groups up to about 60
and decline thereafter (Rönnlund & Nilsson, 2003). Thus, the present pattern may be
part of a more general one in where the point of noticeable decline in fluid aspects of
cognitive functioning occurs considerably later in adulthood than many researchers
conceive of.

The relation between time-related changes in subjective ratings of memory and
changes in memory performance awaits further examination (cf. Jorm, Christensen,
Korten, Jacomb, & Henderson, 2001), but it is interesting to note that mean subjective
ratings of longitudinal change across the first two measurement occasions more
closely adheres to the pattern of the longitudinal than the cross-sectional patterns of
results by showing a trend of accelerated memory changes with advancing age (Study
IV).

Importantly, the cross-sectional data adjusted for differences in quantity of
educational attainment anticipates a similar pattern. Specifically, age no longer
predicted episodic memory performance among the five youngest cohorts once
educational differences were controlled for (Study I and Study IV; but cf. Study III),
suggesting that no substantial changes in average performance levels occurs before age
55 or 60. By contrast, age predicted substantial amounts of variance in memory
performance over and above education across the early born (old) groups, also in line with the longitudinal data. In a similar vein, the cross-sectional and longitudinal aging patterns in semantic memory agreed once differences in education were controlled for (Study IV). Thus, a convergence between the age gradients established by cross-sectional and longitudinal analyses was observed when two potential threats to the internal validity (educational differences and retest effects respectively) were controlled for.

At this particular point it is warranted to reinstate that the relation between education and memory is almost certainly one of mutual influences. In such a case, the decomposition of age-orthogonal and age-related influence by means of the statistical methods used at present (e.g., hierarchical regression) is difficult (Lindenberger & Pötter, 1998). Nevertheless, the observations that education is by far the more important predictor than age across the younger (but not the older) cohorts offer support of the notion that the discrepancy between cross-sectional and longitudinal younger may be the result of cohort-related changes in education. This seems to hold both for episodic and semantic memory. As was noted before matching is the more commonly used means to adjust for educational differences. This method is likely to create differentially representative age-samples. On the other hand, the present use of a statistical control of educational differences is plagued with its own problems. One opportunity offered by virtue of a large data set such as the present one is to examine a sub-set of the total sample that are matched on basis of educational attainment (with regard to quantity and level), and to compare these results with those obtained by statistical adjustment of formal education. In any case, the finding of a relation between cohort differences in memory and cohort differences in education suggest the possibility that human mnemonic capacity may be subject to a relatively rapid “evolution”.

It will be interesting to see whether the present cohort effects reflect the same basic mechanisms as those that underlie generation-wise improvements in other cognitive measures (Flynn, 1987; Neisser, 1998). The present suggestion that the cohort effects are driven by changes within the educational system does not exclude that other cultural factors such as technological advancements, or urbanization (Greenfield, 1998) or improved nutrition (Lynn, 1998) could play a role, of course. Points in time at which cross-generational changes shift in direction or level off are potentially informative regarding the relative importance of these factors. To the extent that mean level differences in body-height are indicative of generational changes in nutrition (e.g., Martorell, 1998) it might be noted that the trend across the samples included in Study IV in body-height is that of a linear improvement across cohorts rather than a positively accelerated increment (primarily observed across the youngest cohorts) in contrast with the data on memory performance and educational attainment.

Turning to a different issue, the age trajectories that emerge from the present studies place some constraints on theoretical accounts proposing that age differences in episodic (and semantic) memory large reflects age-related reduction in one, or a few, basic factors. Specifically, one might expect that age-performance trajectories of the memory measures and measures of the basic factor(s) coincide.

Of interest to note in this regard is that indicators of several of the basic factors that we have considered before appear to be associated with early deficits (i.e., from
age 20 or 30) in contrast with the pattern of a fairly late deficit as established at present. For example, a pattern of a gradual decline from young to old age appears to hold for measures of working memory (Gilinsky & Judd, 1994, Park et al., 2002) and processing speed (Park et al., 2002) and is also discernible across studies neurological factors such as integrity of dopamine functioning (Bäckman et al., 2000; Prull et al., 2000). In fact, a similar trend of an early onset and gradual decline has been observed for gross neuroantomic measures such as brain-weight (e.g., Greenfield et al., 1969).

One obvious possibility is that the apparent discrepancy is due to the fact that negligible losses in these basic factors actually occur before 60. Specifically, the evidence concerning the onset of changes in the potential mediators discussed is still based on cross-sectional age comparisons. It could thus be that the age-related patterns follow the same patterns of changes as the episodic memory measures once cohort-related factors are controlled for. In fact, cohort-related increases in measures such as overall brain-weight have been observed (Miller & Corsellis, 1977) suggesting that the neurological substrate underlying cognition could be subject to short term generation-wise changes (Storfer, 1999) that could bias inferences concerning maturational changes in cross-sectional age-group comparison. The longitudinal evidence concerning the other factors discussed above is also sparse. However, as regards perceptual speed, deterioration appears to begin at an early age also as judged from longitudinal data (Schaie, 1989) suggesting that cohort differences are negligible in this case.

A second possible explanation of the apparent discrepancy between the onset of decline in memory and basic factors is that changes in the latter factors must reach some critical threshold before they have a manifest effect on memory performance. Changes in the dopamine system may be tied to such a threshold effect, at least as regards motor performance (DeCarli, 1995). A similar argument has been raised concerning white matter changes; these must probably reach a critical threshold before behavioral changes are discernible (Boone et al., 1992). The same principle could possibly be advocated to account for a potential discrepancy between longitudinal changes in memory and changes in factors assumed to mediate the age-memory relationships (e.g., processing resources, cf. Park et al., 2002).

A third possibility is that the negative age-related influence exerted by the basic factor(s) from early adulthood is offset by some compensatory mechanism. I have already discussed the possibility that age-related changes in semantic memory measures may be controlled by a negative influence in basic processing efficiency on the one hand and a positive influence in the form of accumulated knowledge on the other. There is much evidence that prior knowledge is an important predictor of episodic remembering (e.g., Bartlett, 1931; Bäckman, 1991; cf. also Tulving, 1985a). It is thus possible that increased knowledge with age also serves to retard the onset of episodic decline to some extent (perhaps more so in everyday settings as novel to-be-remembered information is often part of a more meaning-intense context). If this is so, one might predict that the magnitude of negative age effects is larger for memory of less familiar as compared with more familiar materials. Such a pattern has been observed in some studies (e.g., Knopf, 1991) but is not a universal finding (Erngrund, Mäntylä, & Nilsson, 1996; Kausler, 1994). Turning from a cognitive to a neuroscientific level, the finding of a different pattern of brain activity in older adults,
especially at retrieval, with a bilateral instead of right frontal activation (as is typical for younger adults), has been hypothesized to have a compensatory function (Cabeza, 2002). However, the evidence favoring this hypothesis is mixed and awaits further enquires (Dolcos, Rice, & Cabeza, 2002).

Taken together, the finding of late maintenance of episodic (and semantic) memory performance is intriguing against the background of a early deficits in potential mediating factors. To distinguish between the outlined explanations of the apparent discrepancy between aging patterns in memory and in putative basic factors, further cognitive and neuroscientific research is needed. One important source of evidence regarding the link between changes in memory and potential mediators will be available once the covariance patterns among changes in memory and candidate mediating factors can be analysed. At present, such evidence is sparse (e.g., Sliwinski & Buschke, 1999).

Limitations

Some specific limitations of the present studies are worth mentioning. Although the results obtained in study I and II were based on large and representative samples, the results concerning the influence of retrieval support are somewhat compromised by the fact that the result emanated from a single set of episodes (i.e., two study lists) and the fact that the presentation order among the tests did not vary across the participants. Even though dependencies among tests were examined using parts of the data set there is an obvious confounding of type of test and order/retention interval.

A second limitation regards the markers in Study III and Study IV. In particular, there were only two indicators for the (semantic) knowledge factor. Inclusion of additional markers would have been preferable. The fact that there was a speed demand, albeit, a limited one (the absolute majority of the participants completed all of the items) may have underestimated vocabulary in the oldest cohorts to some minor extent. In addition, a more supportive task in the test of general knowledge (e.g., multiple choice recognition instead of recall would possibly have constituted a more sensitive means to assess knowledge (i.e., by minimizing the demands on fluid abilities/basic resources).

Also, the longitudinal data presented here (Study IV) does not permit firm conclusions regarding the recall versus recognition distinction. Specifically, only one of the markers of recognition used in Study III was included in the longitudinal comparison (Study IV) due to a suspected instrumentation effect on part of the face/name recognition test. However, results by others (Zelinski & Burnight, 1997) indicate very long-term maintenance of recognition even in older adults suggesting that the relatively later maintenance of recognition than recall holds for longitudinal as well as cross-sectional comparisons.

Finally, the test sessions were fairly long. However, potential effects of fatigue were presumably minimized by the large variation in task content. More important, selective effects of fatigue, such that the elderly suffered in particular from this factors should have produced an age by order effect across measures (e.g. within the episodic domain). No such effect is discernible across the measures.
**Future steps**

As noted by Hertzog (1996), cognitive aging research may benefit from starting to test specific hypothesis using extreme age-group designs, then determine the mean level age-trajectories of the cognitive functions of interest, first on basis of cross-sectional data, then using longitudinal analyses. A next important step is to search for predictors of the intra-individual patterns of change. We have already begun to explore some basic issues including the degree of correspondence between changes in levels of cultural/leisure activities and health on the one hand, and changes in memory on the other, bearing on the classic issue in cognitive aging research of whether maintenance of higher activity levels in old age may serve as buffer against cognitive decline (e.g., Hultsch, Hertzog, Small, & Dixon, 1999). A challenging task will be to expand on models including health and life-style factors by adding measures of sensory functions and other biomarkers and to examine the relations to memory factors and other cognitive abilities such as speed. In the 4th data collection wave that has just begun several markers of the latter constructs were added to the Betula battery, which provides a good basis for conducting such analyses in the future.

**Concluding remarks**

The present thesis attempted to shed light on some basic issues in research on memory aging, including the ability structure of declarative memory measures, the mean level changes that are to be expected for these memory abilities, and whether cognitive support may remedy some of the memory losses in old age. Some aspects of the result provide additional support of general trends in the literature including the finding that age-related impairments in old age generalize across cognitive support and differential aging patterns for episodic and semantic memory measures. Others imply that some commonly held views of memory aging may be in need of a revision, including the finding that episodic memory functioning is well preserved until old age. The results raise many new questions that will hopefully be answered by future research efforts.
REFERENCES


Bäckman, L. (1986). Adult age-differences in cross-modal recoding and mental tempo, and older adults utilization of compensatory task conditions. Experimental Aging Research, 12, 135-140.


125-140.


Lichty, W., Bressie, S., & Krell, R. (1988). When a fork is not a fork: Recall of performed activities as a function of age, generation, and bizarreness. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 506-511). Chichester: Wiley.


sectional and 3-year longitudinal memory performance across the adult life span.

Psychology and Aging, 8, 176-186.


Recall of Subject-Performed Tasks, Verbal Tasks, and Cognitive Activities Across the Adult Life Span: Parallel Age-Related Deficits

Michael Rönnlund1, Lars Nyberg1, Lars Bäckman2,3, and Lars-Göran Nilsson4
1Department of Psychology, Umeå University, Umeå, Sweden, 2Department of Psychology, University of Uppsala, Uppsala, Sweden, 3Department of Geriatric Medicine, Karolinska Institute, Stockholm, Sweden, and 4Department of Psychology, University of Stockholm, Stockholm, Sweden

ABSTRACT

Motor activity during encoding of verbal information has been suggested to reduce age differences in episodic memory. Here we examined memory for sentences encoded with enactment (SPTs, subject-performed tasks) or without enactment (VTs, verbal tasks) in a population-based sample consisting of 10 groups ranging in age from 35 to 80 years (N=1000). Memory performance was assessed by immediate free- and category-cued recall. Degree of clustering was measured by the adjusted ratio of clustering score. Recall of cognitive activities served as a complementary measure of memory for performed tasks. Sentence recall showed a gradual decline across age, of about the same magnitude for SPTs and VTs, in both free and cued recall. Clustering in free recall was higher for SPTs than for VTs, but there were no age differences in clustering. A pattern of gradual decline from age 35 was observed also in activity recall, regardless of whether the activities involved motor activity or not. Across the memory measures, differences in education accounted for all of the age-related variance in performance among the younger (35–55 years) but not the older groups (60–80 years), suggesting that genuine aging effects in these measures are more prominent in old age. Together, the results indicate that age differences in episodic memory, in line with most, if not all, types of encoding support, generalize across the performed/non-performed distinction.

Older adults are generally impaired on tests of episodic memory (for reviews, see Craik & Jennings, 1992; Kausler, 1994; Smith, 1996). Age-related deficits in episodic memory are observed across a variety of verbal and non-verbal materials, including concrete and abstract words, sentences and prose, and abstract and concrete pictures (Kausler, 1994; Verhaeghen, Marcoen, & Goosens, 1993). Also, age-related deficits generally persist when encoding is guided by specific instructions. For example, age differences in recall are commonly observed following semantic orienting tasks (Eysenck, 1974) and visual imagery instructions (Dirkx & Craik, 1992). Taken together, these and related findings (e.g. Fisher & McDowd, 1993; Park & Shaw, 1992), indicate that older adults have an impairment in episodic memory that generalizes across various forms of materials and experimental conditions.

One possible exception to this general pattern (see Craik & Jennings, 1992, for further examples) was discovered in the early 1980s when Bäckman and Nilsson (1984, 1985) showed that episodic memory for self-performed actions tended to be age invariant. These studies used an experimental paradigm that had been introduced by several independent researchers a few
years earlier (Cohen, 1981; Engelkamp & Krumnacker, 1980; Saltz & Donnenwerth-Nolan, 1981). In this paradigm, participants are presented with a series of sentences denoting simple actions such as “roll the ball”, “break the match” or “scratch your nose”, for purposes of later recall or recognition. In the experimental condition, the sentences are performed, or enacted, by the participants. Following Cohen (1981), enacted sentences will be referred to as subject-performed tasks (SPTs). In the standard control condition, the sentences are presented without request for enactment (verbal tasks; VTs). A robust finding is that subsequent memory performance is greatly improved by enactment at study (e.g. Bäckman, Nilsson, & Chalom, 1986; Cohen, 1981; Engelkamp & Krumnacker, 1980; Kormi-Nouri, 1995; Nilsson & Craik, 1990; Svensson & Nilsson, 1989; see Cohen, 1989; Engelkamp & Cohen, 1991; Engelkamp & Zimmer, 1994, for reviews). Theoretical accounts of the effect, such as that by Bäckman and Nilsson (1984, 1985), suggest that SPT-recall is superior due to activation of multiple sensory modalities at encoding. The motor component, which may be of central importance in this regard (Engelkamp & Zimmer, 1994; Nyberg et al., 2001), makes this form of encoding support rather unique.

The findings by Bäckman and Nilsson (1984, 1985), that younger and older adults’ memory performance was comparable following enacted encoding, taken to suggest that presence of multimodal cues may offset older adults’ problems of self-initiated recoding processes, sparked much interest, and several follow-up studies were conducted. Some subsequent studies found evidence for a lack of age differences in recall following encoding enactment (Bäckman, 1985a; Brooks & Gardiner, 1994; Dick, Kean, & Sands, 1989), thereby substantiating the initial results. The results of other studies, however, have failed to reproduce this pattern by showing reduced but reliable age differences in recall of enacted as compared with non-enacted sentences (Nyberg, Nilsson, & Bäckman, 1992), or age differences of a similar magnitude regardless of encoding condition (e.g., Cohen, Sandler, & Schroeder, 1987; Guttentag & Hunt, 1988; Knopf, 1991; Knopf & Neidhardt, 1989; Nilsson & Craik, 1990). Although a number of factors have been suggested to account for this mixed picture, no consensus has been reached regarding whether encoding enactment eliminates, reduces, or does not affect age differences in episodic memory. One purpose of the present research was therefore to conduct a large-scale examination of memory for enacted and non-enacted information in adulthood and old age. To this end, cross-sectional data from an ongoing population-based prospective study of memory and health were used (Betula Study; see Nilsson et al., 1997, for a full description of the study). A shortcoming of earlier research is that a select sample of young adults has typically been compared with a sample of old adults. By contrast, the present data set included participants from 10 different age groups in the age range 35–80 years (100 participants per age group). The need to consider a broader age range was recently illustrated by Cregger and Rogers (1998) who suggested that age differences in activity recall may be restricted to differences between young-old (60–70 years) and old-old participants (71–82 years); young (18–34) and young-old subjects did not differ in performance. Another advantage of analyzing data from a large population-based sample is that it gives the possibility to examine the influence of other variables, including demographic (e.g., sex, level of education), social (e.g., active/passive lifestyle), and biological (e.g., health status, genetic composition) factors. In previous analyses of data from the Betula study we have begun to address the effects of multiple-individual-difference variables on memory performance (e.g., Herlitz, Nilsson, & Bäckman, 1997; Nilsson et al., 1996; Nyberg, Bäckman, Erngrund, Olofsson, & Nilsson, 1996).

To keep primary focus on the effect of motor encoding on age-related differences in episodic memory, we presently restrict analyses of background variables to educational attainment. Apart from selection factors, the positive relation between education and episodic memory (e.g., Inouye, Albert, Mohs, Sun, & Berkman, 1993) likely stems from a direct influence of education on cognitive skills, including use of strategies and verbal skills (for a review, see Ceci, 1991). Within a population-based sample, such as the present,
considerable differences in quantity of formal schooling exist among age cohorts, and, thus, education is an important factor to consider when age differences in memory are evaluated.

To determine whether encoding enactment affects the size of age differences in episodic memory performance, several different analyses were conducted. One set of analyses involved comparison of free-recall performance following enacted and non-enacted encoding. Analyses of the free-recall data, presented elsewhere, have indicated a gradual age-related decrease in free-recall performance both for SPTs and VTs (Nilsson et al., 1997; see also Nyberg et al., 1996). However, the foregoing analyses did not directly contrast free-recall scores for SPTs and VTs, leaving open the possibility of differential slopes of the performance curves (cf. Nyberg et al., 1992). Another set of analyses was from a test of cued recall. The nouns in each study list were from different semantic categories. In the cued-recall test, the category labels served as cues for remembering the nouns from the study items (i.e., items encoded as SPTs as well as VTs). Prior evidence suggests that the combination of encoding support in the form of enactment and retrieval support in the form of category cues is particularly helpful for elderly and demented adults (Karlsson, Bäckman, Herlitz, Winblad, & Österlind, 1989). Such findings indicate that even if parallel age effects are observed for SPTs and VTs on the free-recall test, eliminated or reduced age effects may be observed for cued recall of SPTs.

A third set of analyses was performed to investigate the degree of clustering at retrieval. Specifically, adjusted ratio of clustering (ARC-) scores were computed for the free-recall data (Roenker, Thompson, & Brown, 1971). Previous findings suggest that the degree of clustering tends to be higher following enacted encoding for younger as well as older adults (Bäckman & Nilsson, 1984, 1985; Bäckman et al., 1986; but see Engelkamp & Zimmer, 1996). Such findings are of importance for understanding the basis for superior recall of SPTs than VTs in that they indicate that aspects of the SPT-procedure facilitates relational processing (Nyberg, 1993). Moreover, if age differences in free recall of enacted events are observed, analyses of degree of clustering provide information as to whether the age effect reflects age-related differences in the utilization of relational (semantic category) information.

The present study also examined age differences in activity memory. Activity memory was tested by asking participants to recall as many as possible of the cognitive tasks that they had performed during a single test session. Memory for SPTs and experimental activities has been found to be similarly affected by several experimental variables (see Kausler, 1994). The inclusion of an analysis of age differences in activity memory can therefore be seen as a way of increasing the generality of the analyses of age differences in SPT recall. In previous studies of recall of cognitive activities, younger adults have consistently been found to outperform older adults (e.g., Bromley, 1958; Kausler & Hakami, 1983; see Kausler, 1994, for a review). This is in agreement with those studies that have found negative age differences in memory for SPTs, but in conflict with the view that age differences in episodic memory functioning are reduced for study materials involving some kind of performance by the participant.

However, very few studies have examined memory for performed tasks using both recall of enacted sentences and cognitive activities in the same participants (Earles, 1996). Moreover, the age effect in recall of cognitive activities has been found to be accounted for by a subset of the individual test activities included (Kausler & Lichty, 1988; Lichty, Kausler, & Martinez, 1986). In fact, a few activities even seem to show a tendency for a reversed age effect, that is, somewhat higher recall levels for older subjects. For this reason, the present study includes an analysis of overall activity memory as well as memory for different kinds of activities. Specifically, we contrasted memory for activities that could be classified as "non-motor" (e.g., face recognition) with activities that involved a motor component (e.g., Block Design).

METHOD

Participants
One-hundred individuals in each of 10 age cohorts (35, 40, 45, 50, 55, 60, 65, 70, 75, and 80 years) participated. These were recruited through the population registry in
Umeå, a city of about 100,000 inhabitants located in northern Sweden. The sampling was random except that persons with a native tongue other than Swedish, individuals with severe sensory handicaps, organic diseases (e.g., dementia), or mental retardation, were excluded.

Health status and cognitive functioning were assessed on two separate occasions, 1 week apart. In the first session, a trained nurse conducted an extensive examination of health, including blood sampling. Moreover, questionnaires concerning various social variables and life events were administered, and an interview of health status was undertaken. In addition, each participant underwent some testing of cognitive functions. In the second session, a comprehensive battery of cognitive tests was administered by one of three experimenters. Both sessions lasted approximately 2 h. Some select background characteristics are summarized in Table 1. As can be seen, there were approximately 2 h. Some select background characteristics are summarized in Table 1. As can be seen, there were more women (n = 530) than men (n = 470) overall.

Table 1. Background Characteristics Across Age Groups.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Females/males</th>
<th>Years of education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
</tr>
<tr>
<td>35</td>
<td>50/50</td>
<td>13.5</td>
</tr>
<tr>
<td>40</td>
<td>52/48</td>
<td>13.5</td>
</tr>
<tr>
<td>45</td>
<td>54/46</td>
<td>12.7</td>
</tr>
<tr>
<td>50</td>
<td>46/54</td>
<td>10.4</td>
</tr>
<tr>
<td>55</td>
<td>54/46</td>
<td>8.9</td>
</tr>
<tr>
<td>60</td>
<td>46/54</td>
<td>8.8</td>
</tr>
<tr>
<td>65</td>
<td>54/46</td>
<td>8.2</td>
</tr>
<tr>
<td>70</td>
<td>48/52</td>
<td>8.2</td>
</tr>
<tr>
<td>75</td>
<td>53/47</td>
<td>7.5</td>
</tr>
<tr>
<td>80</td>
<td>59/41</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Materials

Enacted/Non-enacted Sentences

Thirty-two sentences in imperative form, each including a unique verb and noun denoting a simple action (e.g., point at the wallet, lift the book), served as the material. The total set of sentences was divided into two lists of 16 sentences. In each list, the nouns of the items belonged to four different semantic categories (four exemplars per category). These were articles of clothing, animals, kitchen utensils, reading materials (first list), parts of the body, musical instruments, carpentry tools and fruits (second list). The exemplars selected were of high to medium typicality in the respective categories, and the degree of typicality did not vary between lists (Nilsson, 1973). Eight random orders of items were prepared for each list. List was counterbalanced across encoding conditions. External objects corresponding to nouns were used for the enactment condition. The sentences to be read were presented in black capital letters printed on white (15 cm × 21 cm) cards.

Activity Memory

The nine activities that served as material for activity recall are listed in Table 2, in the order of presentation. The activities are described in detail in Nilsson et al. (1997). Briefly, the prospective memory task (instruction) requested the participant to remind the experimenter to sign a piece of paper at the end of the test session. In the face/name task (study) the participant silently viewed pictures of children’s faces, together with first names and surnames. The participant was told to try to remember the faces and the surnames. The study trial of SPTs/VTs involved presentation and oral free recall of action commands, that were either acted out motorically by the subject (SPT) or studied verbally (VT), and oral category cued recall of the nouns from both study lists. The fourth activity, name-stem completion, included presentation of two-letter stems (e.g., AN) and the participant was asked to complete each of these with the first surname coming to mind (e.g., ANDERSSON). Fifth, a word fluency task was administered. In each of four subtests of word fluency, the task was to name orally as many words as possible during 1 min, given different restrictions (initial a, initial m and five letters, professions with the initial b, names of animals with initial s including five letters). Next followed a recognition test of faces and names, that included random presentation of faces studied earlier and non-studied faces. The Block Design test (Wechsler, 1981) required the participants to motorically arrange a number (four or nine) of red and white blocks so as to match target patterns depicted on cards under certain time restrictions. In turn, the second test trial of SPT/VT included three subordinate activities: yes/no recognition, followed by cued recall (verb cues) and a source discrimination task. Finally, the participant was presented with four lists of auditorily presented words. These were studied and recalled under four different conditions: focused attention at
study/focused attention at test, divided attention at study/focused attention at test, focused attention at study/divided attention at test, and divided attention at both study and test. Sorting red or white cards into two piles on the basis of color served as the secondary task in the last three conditions.

Procedure

Enacted/Non-enacted Sentences
Each participant was presented with two lists of sentences. The sentences were presented at a rate of 8 s each, and the participant was given explicit instructions to memorize as many sentences as possible. When the last item in each list had been presented, participants were asked to recall orally as many sentences as possible, in any order. In the VT-condition, each sentence was presented visually (cards) and auditorially (read by the experimenter). In the SPT-condition, the participant was required to perform the action described by the sentence. The sentences were read aloud by the experimenter, and for all types of items except those including nouns from the category “body-parts”, an external object to be used in the execution of the command was handed over by the experimenter. The objects were kept in a shelf and were thus only visible during presentation/execution. Two minutes were allowed for free recall of each list. Half of the participants in each age cohort received the VT-condition first, and the other half was given the SPT-condition first.

Following free recall of both lists, participants were given a cued recall task. They were instructed to recall as many as possible of the nouns/objects from the sentences/actions presented earlier. It was emphasized that items from both of the lists were to be recalled. Participants received a sheet of paper including the eight superordinate category names of nouns/objects included in the study lists, and were told that these might serve as a help to remember the nouns/objects. Three minutes were allowed for the cued recall task.

RESULTS

The free- and cued-recall data for enacted and non-enacted sentences will be presented first, followed by the results from the test of activity memory. We finally address the issue of the relative contribution of age and educational differences to memory performance. Only those results significant at an alpha level of 0.01, or better, will be reported, motivated by the large sample size.

Enacted/Non-enacted Sentences

Activity Memory
At the end of the experimental session, participants were instructed to think back and try to recall as many of the tests he/she had performed earlier during the test session (that day). Participants were free to describe the tasks/tests in their own words, but were asked to try to describe them in as much detail as possible. They were allowed to recall the tests in any order and the test was self-paced. No prompts were given. The experimenter marked each identified activity on a protocol that contained a list of the to-be-recalled activities. The experimenters were extensively trained (e.g., they recorded responses together as part of pilot-tests and compared the protocols) to promote high inter-rater reliability. In the few cases where the verbal responses were ambiguous, these were written down and the scoring was made in agreement with the other experimenters.

Free Recall
A strict scoring criterion was adopted in that only responses including the correct verb paired with the correct noun were credited (change in tense
and plural form was allowed). Initial analyses revealed a minor effect of list for both types of encoding (second list easier), but no age × list interaction was observed and we therefore collapsed across this factor. No effect of presentation order was obtained for recall of enacted sentences. However, in the non-enacted condition, recall was somewhat higher when the enacted condition had been presented first (about 7% in terms of mean proportion recalled), indicating some transfer of strategy from enacted to non-enacted encoding. Increased use of motor imagery following enacted encoding (overt enactment was not allowed) is a plausible reason for this effect. As this effect was the same across age we present the data collapsed across presentation order. Reliability estimates (split-half correlations boosted by the Spearman-Brown formula) ranged from .61 to .68 across lists and encoding conditions.

Figure 1 depicts the mean proportion of correctly recalled sentences as a function of age and task. As can be seen, the results indicate a pattern of monotonic decrements in memory performance with advancing age for both types of encoding. Also, levels of recall of enacted sentences were almost twice as high as recall of non-enacted sentences.

A 10 (age group) × 2 (encoding condition: enacted vs. non-enacted) ANOVA, with repeated measures on the last factor supported the foregoing description, in that the main effects of age, $F(9, 990) = 60.12$, $MSE = 8.67$, $\omega^2 = 0.19$, and encoding condition, $F(1, 990) = 1906.98$, $MSE = 3.83$, $\omega^2 = 0.30$, were reliable. These effects showed that performance decreased with increasing age, and that SPTs were better remembered than VTs. The interaction between the two factors was also reliable, $F(9, 990) = 8.03$, $MSE = 3.83$, $\omega^2 = 0.01$, but contrary to what would be expected from the results of the Bäckman and Nilsson (1984, 1985) studies, this was due to a somewhat more pronounced effect of age for enacted items.

Examination of the performance levels indicated that the age × encoding interaction might reflect a subtle floor effect in the oldest cohorts (i.e., in the non-enacted condition). To examine this possibility, a median-split analysis was conducted. The rationale for this analysis is that the restriction in range of scores for the condition in which there is a floor-effect may be remedied by considering the above-median split in that condition. Thus, individuals in each age cohort were divided into two groups: those performing above or below the median in recall of non-enacted sentences (observations at the median were randomly distributed to provide an equal number of persons in each subsample). The outcome of this split is shown in Figure 2A and B.

In line with the overall analysis, an age × encoding interaction characterized the below-median sample, $F(9, 490) = 10.04$, $MSE = 3.14$, $\omega^2 = 0.02$. By contrast, for individuals performing above the median of their age group, the interaction fell short of significance ($F = 1.90$, $p > .05$). Thus, the outcome of these analyses show that the interaction effect apparent in the overall analysis may have reflected a floor effect in the oldest age groups for VTs. Importantly, however, none of the

---

1Recall of verbs scored separately showed a highly similar pattern. Nouns scored separately revealed somewhat attenuated age differences as compared with recall of whole sentences, at least in the enacted condition (a tendency of an age by encoding interaction ($p < .05$) suggestive of larger age differences for SPT recall was still present for nouns). These patterns possibly indicate a problem in specificity of recall of the verbs on part of the elderly. It is possible that the age differences in recall of verbs (and whole sentences) were somewhat exacerbated by the present use of a strict scoring criterion (cf. Earles, Kersten, Turner, & McMullen, 1999).
analyses provided support for the view that enacted information is less susceptible to age differences in free recall than is non-enacted information.

Lack of an age × encoding interaction, or the presence of very weak interaction, might suggest that a substantial portion of the age-related variance in recall of enacted sentences is shared with that of non-enacted sentences. We examined this formally by hierarchic regression analyses, using recall of enacted sentences as the criterion (cf. Earles, 1996). Simple regression analysis indicated that age alone accounted for 36.9% of the variance in recall. Although age still contributed with a significant amount of variance (12.2%) when recall of non-enacted sentences was controlled for by entering this variable before age, the analysis indicated that at least 67% (100–12.2/36.9) of the age-related variance was shared for the two variables. This figure is lower than the corresponding value (80%) reported by Earles (1996), but confirms that a considerable portion of the age-related variance is common for recall of enacted and non-enacted information.

In sum, the free recall data revealed at least as large age-related deficits for enacted as for non-enacted sentences, suggesting that encoding support in the form of enactment does not serve to reduce age differences. Most of the age-related variance in the two conditions was shared, suggesting that recall of enacted and non-enacted tasks draw on common age-sensitive processes.

The ARC measure (Roenker et al., 1971) was used to determine degree of clustering at retrieval. The ARC score of a given protocol is determined by the formula:

$$ARC = \frac{R - E(R)}{\text{Max } R - E(R)}$$

where $R$ represents the summed number of successive repetitions within a given category, $E(R)$ represents the number of repetitions expected to occur by chance, and $\text{Max } R$ is the maximal number of repetitions that can occur. An ARC score of positive unity indicates that the order of responses is completely contingent on categorical membership and one of zero that clustering is at chance level. Calculation of the ARC scores was based on the nouns present in the recall protocols. Table 3 depicts mean ARC scores as a function of age group. A 10 (age) × 2 (encoding condition) ANOVA performed on the ARC score data indicated that age was unrelated to clustering ($p > .20$). The main effect of encoding condition was reliable, $F(9, 990) = 52.96$, $MSE = 0.27$, $\omega^2 = 0.03$, hence replicating the basic finding of Bäckman et al. (1986) that recall of enacted sentences is associated with a higher degree of clustering than recall of non-enacted sentences. The age × task interaction did not approach significance ($p > .30$).

Correlations between recall and ARC-scores were computed. For SPTs as well as VTs there was a weak positive relation, $r = .26$ and .17,
respectively \( r = .22 \) and \( r = .18 \), respectively when age was partialled out).

**Cued Recall**

The total number of nouns recalled when category labels were provided as cues, served as the dependent measure in the analyses of the cued recall data. Figure 3 shows these data across age and encoding condition. As for free recall, a negative age trend is clearly present in cued recall, regardless of encoding condition. A 10 \( \times \) 2 ANOVA showed that the effect of age was highly reliable, \( F(9, 990) = 35.41, \ MSE = 9.99, \ \omega^2 = 0.12 \), as was the effect of encoding condition, \( F(1, 990) = 1521.51, \ MSE = 4.88, \ \omega^2 = 0.29 \). In these analyses, the interaction between the two factors did not approach significance \( (p > .30) \).

In sum, the cued-recall data substantiated the result in free recall by showing (a) an overall advantage of enacted materials and (b) highly similar magnitudes of age-related deficits following enacted and non-enacted encoding. Thus, even when retrieval support is provided, effects of enacted encoding appear to be utilized to about the same extent across age.

**Activity Memory**

The total number of activities recalled served as the dependent measure in the main analysis of activity recall. For those tasks comprising two or several subsets (2, 3, 5, 6, 8 and 9) participants received 1 point for recalling a specific subtest (e.g., “saying words with an initial \( a \)” in test number 5). Recall of the superordinate activity was also credited with 1 point if recalled (e.g. “I was supposed to say words beginning with different letters”). \(^2\) Participants could hence obtain a maximum score of 27 in the activity recall task. The reliability of the test, as estimated by the Spearman–Brown formula, was .78. The mean number of activities recalled across age is shown in Figure 4. A one-way ANOVA showed that

---

### Table 3. ARC Score (M, SDs) Across Age and Encoding Condition.

<table>
<thead>
<tr>
<th>Age</th>
<th>SPTs</th>
<th>VTs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>35</td>
<td>0.32</td>
<td>0.35</td>
</tr>
<tr>
<td>40</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td>45</td>
<td>0.27</td>
<td>0.28</td>
</tr>
<tr>
<td>50</td>
<td>0.29</td>
<td>0.28</td>
</tr>
<tr>
<td>55</td>
<td>0.25</td>
<td>0.31</td>
</tr>
<tr>
<td>60</td>
<td>0.24</td>
<td>0.37</td>
</tr>
<tr>
<td>65</td>
<td>0.21</td>
<td>0.40</td>
</tr>
<tr>
<td>70</td>
<td>0.25</td>
<td>0.38</td>
</tr>
<tr>
<td>75</td>
<td>0.27</td>
<td>0.50</td>
</tr>
<tr>
<td>80</td>
<td>0.17</td>
<td>0.62</td>
</tr>
</tbody>
</table>

---

\(^2\) A significant age-related decrease in recall was evident also at the level of grouped activities (i.e., when recall of any sub-component of the Tasks 1–9 was credited one point).
there was a significant effect of age, \(F(9, 990) = 26.06, \text{MSE} = 8.94, \omega^2 = 0.18\).

Next, we addressed the question of whether different activities exhibit similar age trends with regard to recallability. Towards this end, it was possible to identify activities that involve motor activity and those that do not. If recall of experimental activities is functionally similar to recall of enacted events in the form of SPTs we should predict that: (a) there is a recall advantage for activities including a motor component, and (b) motor and non-motor activities show similar age effects. A comparison of recall levels for the specific subtests that involved a motor component (activities 3a, 7, 9b, 9c, 9d in Table 2) and those that did not (the remaining activities) is provided in Figure 5. The results are consistent with both our hypotheses in that (a) recall levels are higher for motor activities, and (b) there is an age-related decrease in recall levels regardless of motor involvement. An ANOVA confirmed these impressions, yielding main effects of age, \(F(1, 990) = 14.05, \text{MSE} = 0.87\), activity type, \(F(1, 990) = 681.22, \text{MSE} = 16.82\), and no interaction between these two factors, \(F = 1.28, p > .20\).

It could be argued that the activities differ along several dimensions other than motor involvement, and that effects of activity type may be confounded with the effects of input position. We therefore conducted a separate analysis in which recall levels for activities 3a and 9d was compared with those of 3b and 9a. Here, serial position as well as task duration is held constant. In line with the foregoing analysis, we observed main effects of age, \(F(1, 990) = 6.25, \text{MSE} = 1.0\), activity type, \(F(1, 990) = 109.27, \text{MSE} = 4.47\), and no interaction between these factors, \(F = 1.24, p > .20\).

The Relative Contribution of Age and Educational Differences to Memory Performance

So far, the results converge on the notion of a gradual age-related impairment in recall of verbal tasks and performed items from age 35 to 80. To examine the extent to which differences in educational attainment among age groups (Table 1) accounted for the age-related pattern, simple and hierarchic regression analyses were conducted (age entered alone or after education). From Table 1 it can be seen that differences among age cohorts in educational attainment was especially pronounced across the five youngest cohorts (13.9–8.9 years). Thus, differences in education could be of special importance to understand the pattern of age differences in recall across the youngest groups. In order to examine this possibility further, the overall analyses (i.e., including the total age range) were complemented with separate analyses of the five youngest (35–55 years; \(n = 500\)) and the five oldest (60–80 years; \(n = 500\)) cohorts. The results of the regression analyses of five of the recall measures are summarized in Table 4.

The overall analyses (i.e., 35–80) show that age alone accounted for between 15.0 and 34.8% of the variance across the memory measures.\(^3\) When entered after education, age still contributed with a substantial amount of variance. However, the separate analyses of the five youngest and the five oldest age groups yielded a very different pattern of results. Specifically, no unique age-related variance remained when education was entered before age in the analyses involving the younger groups (35–55 years), in contrast with

---

\(^3\) Preliminary analyses indicated that the inclusion of a quadratic age term added 1.2% or less of explained variance across the measures. Therefore only linear effects were considered.
Table 4. Summary of Simple and Hierarchic Regression Analyses of the Recall Measures.

<table>
<thead>
<tr>
<th>Dependent measure</th>
<th>Step</th>
<th>Predictor</th>
<th>Age range</th>
<th>35–80</th>
<th>35–55</th>
<th>60–80</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>35–80</td>
<td>35–55</td>
<td>60–80</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>β</td>
<td>R²</td>
<td>ΔR²</td>
<td>β</td>
</tr>
<tr>
<td>SPT FR</td>
<td>First</td>
<td>Age</td>
<td>-0.590*</td>
<td>.348</td>
<td></td>
<td>-0.223*</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>Education</td>
<td>0.168*</td>
<td>.199</td>
<td>.168*</td>
<td>0.249*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age</td>
<td>-0.495*</td>
<td>.367</td>
<td>.168*</td>
<td>-0.107</td>
</tr>
<tr>
<td>VT FR</td>
<td>First</td>
<td>Age</td>
<td>-0.434*</td>
<td>.188</td>
<td></td>
<td>-0.197*</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>Education</td>
<td>0.282*</td>
<td>.190</td>
<td></td>
<td>0.358*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age</td>
<td>-0.276*</td>
<td>.241</td>
<td>.051*</td>
<td>-0.030</td>
</tr>
<tr>
<td>SPT CR</td>
<td>First</td>
<td>Age</td>
<td>-0.455*</td>
<td>.207</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>Education</td>
<td>0.212*</td>
<td>.161</td>
<td></td>
<td>0.238*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age</td>
<td>-0.336*</td>
<td>.238</td>
<td>.077*</td>
<td>-0.075</td>
</tr>
<tr>
<td>VT CR</td>
<td>First</td>
<td>Age</td>
<td>-0.387*</td>
<td>.150</td>
<td></td>
<td>-0.167*</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>Education</td>
<td>0.243*</td>
<td>.147</td>
<td></td>
<td>0.250*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age</td>
<td>-0.251*</td>
<td>.190</td>
<td>.053*</td>
<td>-0.050</td>
</tr>
<tr>
<td>Activity recall</td>
<td>First</td>
<td>Age</td>
<td>-0.433*</td>
<td>.179</td>
<td></td>
<td>-0.255*</td>
</tr>
<tr>
<td>Recall</td>
<td>Second</td>
<td>Education</td>
<td>0.369*</td>
<td>.241</td>
<td></td>
<td>0.373*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age</td>
<td>-0.216*</td>
<td>.273</td>
<td>.032*</td>
<td>-0.082</td>
</tr>
</tbody>
</table>

*<i>p < .01, F_{change} < 0.01, FR = free recall, CR = cued recall.</i>
the analyses involving the older groups (60–80 years). Thus, most of the differences in recall among the younger, but not the older groups, may be attributable to differences in educational attainment.

DISCUSSION

The main objective of the present study was to examine recall performance across the adult life span for events that did or did not involve enactment during study. To this end, we examined free and cued recall of enacted and non-enacted sentences, as well as recall of complex cognitive tasks (activity memory). For both types of enacted events, we observed a robust and gradual deterioration of performance with increasing age.

Starting with the activity memory task, the results are in line with previous findings of reliable age differences in recall (e.g., Earles & Kersten, 1998; Kausler, 1994). Analyses of recall of individual activities revealed that activities that involved overt performance were better remembered than activities that did not involve motor activity. Most importantly, age was negatively associated with performance for motor as well as non-motor activities (cf. Lichty et al., 1986), in agreement with the results from the comparison of memory for enacted and non-enacted sentences. Specifically, the latter results indicated (a) a continuous decrease in free recall performance across the adult life span following both types of encoding, and (b) a substantial recall advantage of enacted over non-enacted sentences. Thus, although enactment facilitated recall performance, the magnitude of facilitation was about the same in all age groups.

The apparent discrepancy between the present finding of a continuous deterioration in performance, and the results by Cregger and Rogers (1998), that were suggestive of an age-related impairment only for the older adults, is possibly resolved by considering a difference between studies with regard to sample composition. Specifically, the study by Cregger and Rogers involved age groups with comparable educational background, whereas groups in the present study differed considerably in educational attainment. In support of the view that the apparent early onset of decline may be attributable to cohort-differences in education, our regression analyses showed that age no longer predicted recall across the age range 35–55 years when years of education was partialed out. This held true for recall of activities as well as for recall of enacted/non-enacted sentences. By contrast, a substantial portion of unique age-related variance remained for the older half of the sample (60–80 years). Thus, true age-related deficits may be more prominent in the period from young-old to old-old age, in line with the results by Cregger and Rogers (1998).

Analyses of categorical clustering revealed no age differences. This finding is consistent with the results from several previous studies (e.g. Bäckman & Larsson, 1992; Brown-Whistler & Freund, 1993), indicating that age differences in free recall of enacted events cannot be attributed to age differences in utilization of relational (semantic category) processing (cf. Bäckman & Nilsson, 1984, 1985). Our finding of a main effect of encoding condition on degree of clustering suggests that the recall advantage of SPTs over VTs may not exclusively reflect increased item-specific processing in the case of SPTs (Olofsson, 1997; Zimmer & Engelkamp, 1989), but also increased relational information (cf. Nyberg, 1993). As noted by Nyberg (1993, p. 31), it is possible that the presence of real objects rather than enactment per se is crucial for enhancing encoding of commonalities among the input events. It is important to note, however, that whereas level of clustering was low in the verbal condition in the present experimental situation (immediate recall of short lists of sentences with no guided encoding instructions), it may be higher under other conditions. For example, if the encoding instruction had mentioned that items could be grouped, retrieval of verbal tasks would likely have been guided by category type to a greater extent. However, with non-guided instructions, semantic organization seems more prominent in the enacted condition. In addition, the analyses of cued-recall performance revealed no tendency for an age by encoding interaction. This observation adds to the free-recall data, indicating that the
magnitude of age differences in recall remains unchanged when support is provided at both encoding (enactment) and retrieval (category cues; see Erngrund, Mäntylä, & Rönnlund, 1996 for a similar finding using verb cues).

Taken together, these data indicate that age differences in recall performance are equally large for enacted and non-enacted events. As noted by previous reviewers, it is difficult to determine the reasons for the discrepant results in published studies, ranging from lack of age differences to quite pronounced negative age effects in SPT recall (Craik & Jennings, 1992; Nilsson & Craik, 1990; Nyberg et al., 1992).

Table 5 summarizes the result obtained in 29 studies/experiments conducted to date. The studies are listed according to the principle findings. These were indicative either of (a) parallel age effects for enacted and non-enacted conditions (1–14), (b) reduced age effects following enactment (15–21), (c) parallel and reduced effects according to the level of a third (moderator) variable (22–23), or (d) age effects for enacted material in the absence of a control task (24–29).

Several observations are noteworthy. First, all studies report a numerical advantage of younger over older adults. Second, studies consistent with parallel age effects (n = 14) outnumber those indicating reduced effects for SPTs (n = 7). Third, the result of parallel age effects has been found to generalize across a number of additional experimental variables. These include levels-of-processing (Nilsson & Craik, 1990), degree of familiarity of the actions (Knopf, 1991), command complexity (Knopf, 1995b), presence or absence of real-life objects (Knopf, 1992, Exp. 4), and retention interval (Knopf, 1992; Knopf & Neidhardt, 1989). Thus, the finding of parallel age differences for enacted and non-enacted material in the absence of a control task (24–29).

In summary, a single factor that account for the mixed outcomes as regards age effects for recall of enacted and non-enacted information is not apparent. However, a possible hint to a boundary condition for the finding of reduced age effects was provided by Knopf (1995a). When comparing her results with those of Bäckman and Nilsson (1984), she noted that the relative effect of enactment differs between studies. Specifically, in contrast with her data and those typically reported in young adult studies, there was essentially no enactment effect for the younger participants in the Bäckman and Nilsson (1984) study. Interestingly, results from the studies reporting reduced and parallel effects within the same study (Brooks & Gardiner, 1994, Exp. 2; McDonald-Miszczak et al., 1996) also conform to this pattern, in that there is essentially no SPT effect for the young in the condition where age effects for SPTs are reduced. In fact, only two experiments reported reduced age effects together with a significant SPT effect in the young (Dick et al., 1989, Exp. 2; Nyberg et al., 1992), and the reduction of the age effect in these studies was relatively modest.
<table>
<thead>
<tr>
<th>Study/experiment</th>
<th>Age Y/N</th>
<th>Age O/N</th>
<th>Enc manip</th>
<th>Lists/cond</th>
<th>List length</th>
<th>List type</th>
<th>Prec rate</th>
<th>Test</th>
<th>Condition</th>
<th>Control task</th>
<th>SPT-effect Y</th>
<th>O</th>
<th>Age-effect</th>
<th>Reduced age effect?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nilsson &amp; Craik (1990) – 1</td>
<td>20.7</td>
<td>68.9</td>
<td>Within</td>
<td>1</td>
<td>36</td>
<td>Concrete</td>
<td>8</td>
<td>DFR</td>
<td>Standard</td>
<td>VTs</td>
<td>42</td>
<td>27</td>
<td>13</td>
<td>No</td>
</tr>
<tr>
<td>2. Knopf (1992) – 4</td>
<td>25.7</td>
<td>69.0</td>
<td>Within</td>
<td>*</td>
<td>16</td>
<td>Concrete</td>
<td>*</td>
<td>IFR</td>
<td>Concrete</td>
<td>VTs</td>
<td>30</td>
<td>24</td>
<td>–</td>
<td>8</td>
</tr>
<tr>
<td>3. Cohen et al. (1987) – 1</td>
<td>21.3</td>
<td>69.4</td>
<td>Within</td>
<td>1</td>
<td>37</td>
<td>Concrete</td>
<td>5</td>
<td>IFR</td>
<td>Standard</td>
<td>Words</td>
<td>24</td>
<td>18</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>5. Knopf (1992) – 3</td>
<td>25.0</td>
<td>61.2</td>
<td>Within</td>
<td>*</td>
<td>12/36</td>
<td>Symbolic</td>
<td>*</td>
<td>IFR</td>
<td>Short list</td>
<td>VTs</td>
<td>21</td>
<td>14</td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>6. Earles (1996)</td>
<td>20.3</td>
<td>67.2</td>
<td>Within</td>
<td>2</td>
<td>12</td>
<td>Concrete</td>
<td>5</td>
<td>IFR</td>
<td>Standard</td>
<td>VTs</td>
<td>20</td>
<td>17</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>7. Knopf (1992) – 2</td>
<td>25.7</td>
<td>63.4</td>
<td>Within</td>
<td>*</td>
<td>16</td>
<td>Symbolic</td>
<td>*</td>
<td>IFR</td>
<td>Standard</td>
<td>VTs</td>
<td>15</td>
<td>15</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>8. Cohen et al. (1987) – 2</td>
<td>20.8</td>
<td>67.8</td>
<td>Within</td>
<td>1</td>
<td>15/27</td>
<td>Concrete</td>
<td>5</td>
<td>IFR</td>
<td>Short list</td>
<td>VTs</td>
<td>15</td>
<td>20</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>9. Guttentag &amp; Hunt (1988)</td>
<td>18.7</td>
<td>74.5</td>
<td>Within</td>
<td>1</td>
<td>24</td>
<td>Concrete</td>
<td>10</td>
<td>IFR</td>
<td>Standard</td>
<td>VTs+imag</td>
<td>13</td>
<td>8</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>10. Norris &amp; West (1993)</td>
<td>19.7</td>
<td>68.7</td>
<td>Between</td>
<td>2</td>
<td>16</td>
<td>Obj free</td>
<td>6</td>
<td>IFR</td>
<td>Verb test/org</td>
<td>VTs</td>
<td>12</td>
<td>16</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>11. Norris &amp; West (1991)</td>
<td>19.7</td>
<td>68.7</td>
<td>Within</td>
<td>1</td>
<td>6</td>
<td>Concrete</td>
<td>8/45</td>
<td>FR</td>
<td>No obj/fast</td>
<td>Cogn activ</td>
<td>10</td>
<td>25</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>12. Knopf (1995a, 1995b)</td>
<td>22.6</td>
<td>69.8</td>
<td>Within</td>
<td>1</td>
<td>22</td>
<td>Symbolic</td>
<td>10</td>
<td>IFR</td>
<td>Simple</td>
<td>VTs</td>
<td>6</td>
<td>3</td>
<td>15</td>
<td>18</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study/experiment</th>
<th>Age Y/N</th>
<th>Age O/N</th>
<th>Enc manip With/betw</th>
<th>Lists/ cond</th>
<th>List length</th>
<th>Enact type conc/symb</th>
<th>Pres rate</th>
<th>Test</th>
<th>Condition</th>
<th>Control task</th>
<th>SPT-effect</th>
<th>Age-effect</th>
<th>Reduced age effect?</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Nilsson &amp; Craik (1990) – 2</td>
<td>18.9</td>
<td>71.4</td>
<td>Between</td>
<td>6</td>
<td>12</td>
<td>Concrete</td>
<td>5</td>
<td>IFR</td>
<td>Deep enc</td>
<td>VTs</td>
<td>4</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>14. Knopf (1991)</td>
<td>24.3</td>
<td>64.9</td>
<td>Within</td>
<td>3</td>
<td>22</td>
<td>Symbolic</td>
<td>13</td>
<td>IFR</td>
<td>High fam</td>
<td>VTs</td>
<td>1</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>15. Dick et al. (1989) – 2</td>
<td>21.2</td>
<td>73.2</td>
<td>Within</td>
<td>1</td>
<td>32</td>
<td>Concrete</td>
<td>Self-paced</td>
<td>FR</td>
<td>Deep enc</td>
<td>EPTs</td>
<td>19</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>16. Nyberg et al. (1992)</td>
<td>19.1</td>
<td>69.3</td>
<td>Within</td>
<td>6</td>
<td>12</td>
<td>Concrete</td>
<td>5</td>
<td>IFR</td>
<td>Standard</td>
<td>VTs</td>
<td>20</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>17. Bäckman &amp; Nilsson (1984)</td>
<td>20.7</td>
<td>71.4</td>
<td>Between</td>
<td>7</td>
<td>12</td>
<td>Concrete</td>
<td>5</td>
<td>IFR</td>
<td>Standard</td>
<td>VTs</td>
<td>11</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>18. Bäckman &amp; Nilsson (1985)</td>
<td>20.4</td>
<td>72.6</td>
<td>Between</td>
<td>8</td>
<td>12</td>
<td>Concrete</td>
<td>5</td>
<td>IFR</td>
<td>Standard</td>
<td>VTs</td>
<td>13</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>19. Brooks &amp; Gardiner (1994) – 1</td>
<td>20.0</td>
<td>69.0</td>
<td>Within</td>
<td>3</td>
<td>15</td>
<td>Symbolic</td>
<td>6</td>
<td>IFR</td>
<td>Standard</td>
<td>VTs</td>
<td>8</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>20. Dick et al. (1989) – 1</td>
<td>22.7</td>
<td>71.2</td>
<td>Within</td>
<td>2</td>
<td>12</td>
<td>Concrete</td>
<td>5</td>
<td>FR</td>
<td>Standard</td>
<td>VTs</td>
<td>3</td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td>21. Bäckman (1985a)</td>
<td>22.5</td>
<td>69.8</td>
<td>Between</td>
<td>6</td>
<td>12</td>
<td>Concrete</td>
<td>5</td>
<td>IFR/DFR</td>
<td>Standard</td>
<td>VTs + imag</td>
<td>3</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>22. MacDonald- Miszczak et al. (1996)</td>
<td>21.4</td>
<td>71.6</td>
<td>Between</td>
<td>1</td>
<td>35</td>
<td>Concrete</td>
<td>6–10</td>
<td>IFR/DFR</td>
<td>Trial 1</td>
<td>VTs</td>
<td>9</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>23. Brooks &amp; Gardiner (1994) – 2</td>
<td>25.0</td>
<td>68.0</td>
<td>Within</td>
<td>2</td>
<td>12/24</td>
<td>Symbolic</td>
<td>6</td>
<td>IFR</td>
<td>Short list</td>
<td>VTs</td>
<td>9</td>
<td>20</td>
<td>13</td>
</tr>
<tr>
<td>24. Lichty, Bressie, &amp; Krell (1988) – 2</td>
<td>22.0</td>
<td>69.0</td>
<td>Generated</td>
<td>1</td>
<td>24</td>
<td>Concrete</td>
<td>10</td>
<td>FR</td>
<td>Given</td>
<td>VTs</td>
<td>17</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Study</td>
<td>Authors</td>
<td>Encoding</td>
<td>Manipulation</td>
<td>Between participants</td>
<td>Concrete</td>
<td>Type</td>
<td>Control task</td>
<td>Interference</td>
<td>Immediate Free Recall</td>
<td>Delayed Free Recall</td>
<td>Familiarity</td>
<td>Familiarity Type</td>
<td>Cogn. Activ.</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>----------</td>
<td>--------------</td>
<td>---------------------</td>
<td>----------</td>
<td>------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>------------</td>
</tr>
<tr>
<td>25.</td>
<td>Kausler, Wiley, &amp; Phillips (1990)</td>
<td>18.9</td>
<td>71.4</td>
<td>–</td>
<td>1</td>
<td>16</td>
<td>Concrete</td>
<td>5</td>
<td>DFR</td>
<td>1 pres</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 pres/massed</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 pres/distrib</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>26.</td>
<td>Phillips &amp; Kausler (1992)</td>
<td>21.4</td>
<td>75.2</td>
<td>–</td>
<td>1</td>
<td>12</td>
<td>Concrete</td>
<td>*</td>
<td>DFR, 10 min</td>
<td>Same room</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diff room</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DFR, 24 h</td>
<td>Same room</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diff room</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>27.</td>
<td>Kausler, Wiley, &amp; Lieberwitz (1992)</td>
<td>20.3</td>
<td>71.9</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>Concrete</td>
<td>–</td>
<td>DFR</td>
<td>No interf</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>72</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Verb interf</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EPT interf</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SPT interf</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>28.</td>
<td>Wiley &amp; Kausler (1993)</td>
<td>21.3</td>
<td>71.0</td>
<td>–</td>
<td>1</td>
<td>24</td>
<td>Concrete</td>
<td>4</td>
<td>DFR</td>
<td>Intentional</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Incidental</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>29.</td>
<td>Lichty et al. (1988)</td>
<td>20.0</td>
<td>70.0</td>
<td>–</td>
<td>1</td>
<td>24</td>
<td>Concrete</td>
<td>10</td>
<td>FR</td>
<td>Given/ordi</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Given/unus</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gener/ordi</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gener/unus</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Note.** enc = encoding; manip = manipulation; with = within participants; betw = between participants; enact = enactment; conc = concrete; symb = symbolic; pres = presentation; ctrl = control task; IFR = immediate free recall; DFR = delayed free recall; * = unspecified; med = medium; fam = familiarity; imag = imagery instruction; verb = verbal; obj = object; org = organized list; unor = unorganized list; cogn act = cognitive activities; EPTs = experimenter-performed tasks; mot = motor test; distrib = distributed; diff = different; interf = interference; ordi = ordinary action; unus = unusual action; gener = generated.
participants and the difference in size of the age effect for VTs and SPTs.

Hence, the critical factor here may be younger participants’ gain from this form of encoding support, rather than whether older participants perform at a similar level as younger adults following encoding enactment. Under certain conditions, younger adults’ ability to spontaneously engage in self-initiated encoding processes (Bäckman, 1985b, 1989; Craik, 1983, 1986) may leave little room for further improvement by provision of support in the form of enactment. In such cases, which may be characterized by high performing younger participants, short lists, multiple trials, organizable items, and so forth, or some combination of these factors, the relative gain from encoding enactment will be greater for older than for younger participants.

Based on the present results, and the results summarized in Table 5, however, we suggest that the typical outcome is that both younger and older adults benefit from encoding enactment, and hence that an age by task interaction in recall of enacted and non-enacted events constitutes an exception to the rule. Recent results from the Betula study show that the pattern of similar age-related differences for enacted and non-enacted information is observed also in recognition (Lövden, Rönnlund, & Nilsson, in press; see also Knopf, 1991). Thus, the pattern of age differences in memory for enacted and non-enacted events can be arranged along with the pattern seen for most experimental manipulations in research on aging and episodic memory. The core feature of the database in this area is that sensitivity to variations in encoding instructions, retrieval support, and the materials, varies little from early through late adulthood. When age-related variations occur (i.e., when age by condition interactions are present), this often seem reflect factors related to the scale of measurement, procedural constraints, and so forth, rather than genuine age changes in the processes under scrutiny (Bäckman, Small, Wahlin, & Larsson, 1999; Craik & Jennings, 1992; Verhaegen et al., 1993). The notion that younger and older adults are equally sensitive to various forms of encoding and retrieval support has implications for the conceptualization of age-related differences in episodic memory. This view seems to imply that the basis for age deficits cannot be fully explained in terms of deficiencies in specific encoding or retrieval operations. Such deficiencies likely reflect more general age-related changes, such as neural changes (e.g., Bäckman et al., 1997; Cabeza et al., 1997; Nyberg et al., 1996) or changes in speed and/or sensory or cognitive capacities (e.g., Hultsch, Hertzog, Dixon, & Small, 1998; Lindenberger & Baltes, 1997; Salthouse, 1996). Anderson and Craik (2000) recently proposed a cascade of age-related neurological and cognitive changes that may lead to memory impairments in older adults. By their view, neurological changes in specific brain regions lead to reduced attentional resources and cognitive slowing. In turn, this causes a reduction in cognitive control, that is the ability to manage one’s thoughts, recollections, and actions in accordance with task-relevant goals. This reduction in cognitive control, finally, is supposed to negatively affect many forms of memory, including episodic memory.

The foregoing view may appear very pessimistic in that memory deficits in old age seem unavoidable. However, as emphasized by Anderson and Craik (2000), and as evidenced by the present data, various forms of support seem able to functionally alter the cascade of neurocognitive aging. For example, Logan, Sanders, Snyder, Morris, and Buckner (2002) recently demonstrated that the lower level of left prefrontal activity during encoding typically observed in older compared to younger adults (see Cabeza & Nyberg, 2000) can be increased when appropriate guidance is provided at encoding. In parallel with this increase, improved subsequent memory performance was observed (Logan et al., 2002). Thus, a more fruitful strategy for future studies in cognitive aging may be to investigate conditions that lead to maximal performance enhancement for the elderly, and not be too concerned with whether the enhancement is of equal or different magnitude than that for younger adults.

CONCLUSIONS

The present results indicate that age-related differences in episodic memory are of similar
magnitude for events that involve performance on part of the participant and those that do not. This is consistent with the observation that age differences generalizes across a variety of materials and levels of cognitive support (Verhaeghen et al., 1993). The finding that a substantial portion of the age-related variance in memory for performed items is shared with that of non-performed items (Earles, 1996) suggests that common age sensitive processes are involved regardless of level of encoding support. The present observations of an age-related deficit in recall of cognitive activities, irrespective of motor involvement, provide further support for the robustness and general nature of age-related episodic memory deficits.

ACKNOWLEDGEMENTS

The present research was supported by grants to Lars-Göran Nilsson from the bank of Sweden Tercentenary Foundation, the Swedish Council for Planning and Coordination of Research, the Swedish Council for Research in the Humanities and Social Sciences, and the Swedish Council for Social Research. We thank Gunilla Smedberg-Åman and Anna-Lisa Nilsson for their assistance in collecting the memory data.

REFERENCES


Remembering and Knowing in Adulthood: Effects of Enacted Encoding and Relations to Processing Speed

Martin Lövdén¹, Michael Rönnlund², and Lars-Göran Nilsson¹
¹Department of Psychology, Stockholm University, Stockholm, Sweden, and ²Department of Psychology, Umeå

ABSTRACT

We examined age differences in recollective experience accompanying recognition using Tulving’s (1985) remember/know procedure, in a population-based sample of adults ranging in age from 35 to 90 years (N = 323). A guided encoding condition (enactment) was compared with a non-guided encoding condition. The relation between age differences in remembering and measures of processing speed was furthermore examined. The results demonstrated higher levels of remembering following enacted encoding, but age-related decrease in remember reports following both non-guided and guided encoding. The frequency of know-reports exhibited an age-related decrease, but this effect was weak. Importantly, a large amount of the explained variance in remembering was shared by age and processing speed. Together, these results are more consistent with a common cause or processing deficiency interpretation, than with a production deficiency interpretation of age differences in recollective experience.

Age-related differences in episodic memory performance are well documented (see Kausler, 1994 for review). In addition to the quantitative decline depicted by these results, recent evidence suggests that advanced age is associated with experiential changes (e.g., Parkin & Walter, 1992; Perfect & Dasgupta, 1997). Tulving (1985) introduced an influential approach to the study of phenomenal aspects of memory retrieval, the Remember/Know procedure. This procedure, typically implemented in a recognition test, requires participants to report whether they consciously recollect an item’s prior occurrence or whether they just know it was there. Remembering (R) is reported when recognition is accompanied by retrieval of any contextual aspect of the studied event, including internally or externally derived information (e.g., thoughts, images, color, or sounds). Otherwise, knowing (K) is reported.

A wealth of theoretical accounts of the hypothetical structures/processes underlying the R/K-reports have been put forward (see Gardiner, Ramponi, & Richardson-Klavehn, 1998 for review). Originally, the R/K procedure was introduced to measure autonoetic and noetic states-of-awareness, assumed to reflect episodic and semantic memory, respectively (Tulving, 1985). Other theories conceive of remembering and knowing as differing in involvement of recollection and familiarity (e.g., Jacoby, Yonelinas, & Jennings, 1997; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998), or as experiential states relying upon distinctiveness and fluency (Rajaram, 1996, 1998). Still others have proposed single process models, suggesting that remember and know judgements map onto a continuum of memory strength (Donaldson, 1996; Hirshman, 1998; Hirshman & Master, 1997).

Address correspondence to: Martin Lövdén, Department of Psychology, Stockholm University, S-106 91 Stockholm, Sweden. E-mail: mal@psychology.su.se
Accepted for publication: May 13, 2002.
Table 1. Summary of R/K-studies Comparing Younger and Older Adults in Recognition Memory.

<table>
<thead>
<tr>
<th>Study</th>
<th>Young-old difference</th>
<th>Encoding Task</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Remember</td>
<td>Know</td>
</tr>
<tr>
<td>Fell (1992)</td>
<td>−.01</td>
<td>.02</td>
</tr>
<tr>
<td>Mark &amp; Rugg (1998)</td>
<td>−.00</td>
<td>−.02</td>
</tr>
<tr>
<td>Schacter et al. (1997; Exp. 1)</td>
<td>.11</td>
<td>−.01</td>
</tr>
<tr>
<td>Perfect et al. (1995; Exp. 2a)</td>
<td>−.01</td>
<td>.21*</td>
</tr>
<tr>
<td>Perfect et al. (1995; Exp. 2a)</td>
<td>.06</td>
<td>.16*</td>
</tr>
<tr>
<td>Schacter et al. (1997; Exp. 2)</td>
<td>.19*</td>
<td>.07</td>
</tr>
<tr>
<td>Fell (1992)(^a)</td>
<td>.33*</td>
<td>−.36*</td>
</tr>
<tr>
<td>Parkin &amp; Walter (1992; Exp. 1)</td>
<td>.32*</td>
<td>−.21*</td>
</tr>
<tr>
<td>Parkin &amp; Walter (1992; Exp. 2)</td>
<td>.28*</td>
<td>−.15*</td>
</tr>
<tr>
<td>Perfect et al. (1995; Exp. 1)</td>
<td>.35*</td>
<td>−.30*</td>
</tr>
<tr>
<td>Perfect &amp; Dasgupta (1997)(^b)</td>
<td>.23*</td>
<td>−.05</td>
</tr>
<tr>
<td>Perfect &amp; Dasgupta (1997)(^c)</td>
<td>.30*</td>
<td>−.06</td>
</tr>
<tr>
<td>Norman &amp; Schacter (1997)</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>Trott et al. (1999)</td>
<td>.07</td>
<td>−.01</td>
</tr>
</tbody>
</table>

Note. Positive difference scores indicate lower hit/target mean rates for old as compared to young groups: \(^a\)From the repetition condition. \(^b\)From the word condition. \(^c\)From the nonword condition. *Indicates significant age-differences.

The present study does not intend to resolve this major theoretical issue, but to focus on age differences in recollective experience accompanying recognition from the point of the R/K procedure. Several prior studies have approached this issue. The results are summarized in Table 1. To facilitate comparison across studies, the outcome of each study is presented in the form of a young-old difference score (The mean level of R/K-judgements reported was first divided by the total number of targets. The mean level of R/K-judgements for the older participants was then subtracted from the mean level of the younger participants). For reasons that will be evident below, the nature of the encoding task (guided versus non-guided) is indicated.

As concerns the relative frequency of remember experiences, 7 out of 14 conditions demonstrated a significant young-old difference, consistent with lower levels for the old. In some conditions a tendency in the same direction is discernable, with a mean difference across conditions of .16. However, in other cases the level of remember-experiences was virtually identical for young and old participants. Thus, although there is a decided trend towards an age-related decrease in remembering, there are exceptions. As regards knowing, the results reveal a mixed pattern of findings. In eight studies no age differences were observed, whereas in four conditions K-reports were higher for the older participants. Moreover, the data in two conditions (Perfect, Williams, & Anderton-Brown, 1995) are consistent with an age-related decrement in knowing.

Concerning episodic memory performance, age-related deficits have frequently been attributed to encoding problems on behalf of elderly (see Kausler, 1994 for review). Specifically, it has been proposed that older individuals tend to engage in stereotypic and schematic processing, at the expense of encoding of idiosyncratic and contextual information (see Burke & Light, 1981; Light, 1991 for reviews). Remember experiences are, by definition, reliant on successful retrieval of contextual aspects of the study event, and, consequently, impaired encoding has been proposed as a main determinant of the age differences in recollective experience as well (Mäntylä, 1993; Perfect & Dasgupta, 1997). In line with this notion, Mäntylä found that stereotypic encoding characterized older adults' encoding strategies (using a cued-recall paradigm), and that the nature of encoding was related to the diminished levels of R-reports in old participants. In a similar vein, Perfect and Dasgupta found that items encoded elaborately, as indexed by verbal reports, were as likely to be
remembered by old as by young participants. Moreover, age differences in response bias do not appear to be a likely explanation of age differences in remembering as confidence ratings for the judgements do not vary across age (Perfect et al., 1995). Thus, several lines of evidence suggest that age-deficits in reported recollective experience are related to encoding problems on behalf of older adults.

A different issue is whether inclusion of specific instructions or materials at the time of encoding may serve to compensate for such deficits. A possibility is that instructions or materials that guide encoding towards further analysis of conceptual or sensory characteristics of the to-be-remembered material help older individuals overcome problems of spontaneously initiating effective encoding strategies, thereby minimizing the age-related memory deficits (e.g., Bäckman, 1985, 1989). The evidence favoring such a production deficiency hypothesis is less than overwhelming, at least when overall memory performance is considered. This is because many studies have failed to detect an Age × Encoding Task interaction. Some studies have even found that younger individuals benefit more than older ones from provision of instructions or materials that serve to facilitate encoding (Burke & Light, 1981; Craik & Jennings, 1992 for reviews). Nevertheless, it has been suggested that discrepancies across R/K studies may be understood from considering the guided/non-guided encoding distinction (Perfect & Dasgupta, 1997; Perfect et al., 1995).

As previously noted, and discernible from Table 1, K-report levels are higher for the older adults in some, but not all, studies. Perfect et al. (1995; see also Perfect & Dasgupta, 1997) hypothesized that with non-guided encoding, maintenance rehearsal may increase with advancing age at the expense of attention to distinctiveness and context. This should result in an age-related increment in K-judgements in non-guided conditions, since maintenance rehearsal magnifies levels of knowing (Gardiner, Gawlik, & Richardson-Klavehn, 1994). Provided that encoding is guided, the age differences in maintenance rehearsal may, on the other hand, be attenuated and consequently result in reduced age differences in knowing. As regards remembering, a similar picture might be expected, that is reduced age-effects given that encoding was guided. The results in Table 1 offer some support for such a production deficiency interpretation. Specifically, findings of lack of age differences in remembering primarily emanate from conditions in which encoding was guided (Table 1, upper part), whereas age differences are more commonly found in non-guided conditions (Table 1, lower part). In addition, magnified K-levels for the older participants tend to be observed in non-guided conditions. The generalization of this pattern, however, is complicated by findings of a young-old difference in remembering in some guided conditions (Schacter, Koutstaal, Johnson, Gross, & Angell, 1997), and lack of such a difference in some non-guided conditions (Trott, Friedman, Ritter, Fabiani, & Snodgrass, 1999). It is finally to be noted that the pattern consistent with a production deficiency hypothesis is emergent from comparisons across rather than within studies. Clearly, further work is needed to examine the extent to which age differences in recollective experience are controlled by the nature of the encoding conditions.

A main objective of the present study was to provide further evidence concerning this issue. To this end, we compared a non-guided instruction to encode simple sentences (e.g., roll the ball), with a guided encoding condition in which participants enacted the sentences. The enactment effect, that is the mnemonic advantage of enacted over non-enacted information, is highly reliable. This effect is typically regarded to reflect enhanced item-specific processing associated with performing the actions, related to assimilation of motor/multi modal information (Bäckman & Nilsson, 1984, 1985; Engelkamp & Zimmer, 1984; Heil et al., 1999; see Engelkamp, 1998; Nilsson, 2000 for reviews). Importantly, two young adult studies observed large advantages of enacted over non-enacted encoding as reflected by remember experiences (Conway & Dewhurst, 1995; Engelkamp & Dehn, 1997), suggesting that the information stored as a consequence of enactment is very effective in triggering subsequent remember experiences.

A number of studies have compared enacted and nonenacted encoding using overall performance measures. With respect to recall, the more general outcome seems to be that of equal age
differences for enacted and nonenacted material (e.g., Cohen, Sandler, & Schroeder, 1987; Earles, 1996; Nilsson & Craik, 1990; Rönnlund, Nyberg, Bäckman, & Nilsson, 2002; but see Bäckman & Nilsson, 1984, 1985). As concerns recognition measures, relatively few studies are available. Two studies found reduced age differences following enacted encoding (e.g., Knopf & Neidhardt, 1989; Nilsson & Craik, 1990), but the results were complicated by near ceiling level performance in the enacted condition. Other studies found age differences of similar magnitude following enacted and nonenacted encoding (Knopf, 1991; Larsson & Bäckman, 1994), suggesting that overall recognition performance might be expected to reveal age differences following enacted, as well as nonenacted encoding. It is important to note that an age by encoding condition interaction may emerge when recognition is decomposed into remembering and knowing, even though overall performance data are suggestive of parallel age-related differences across encoding conditions. Specifically, if older participants’ recognition performances are largely based on knowing when encoding is non-guided (e.g., Parkin & Walter, 1992), and if guided instructions ensure that the proportion of R-reports are high across age, purifying the memory measure by exclusively considering R-reports might accentuate age-related differences in non-guided conditions. Thus, in line with the production deficiency hypothesis outlined above, an age by encoding interaction may emerge when recognition is decomposed into R/K experiences.

Whereas it is clearly important to control for specific factors, such as those that co-vary with encoding conditions, it may in addition be important to examine the potential influence of more general subject-related factors to explain the age-related effects in recollective experience. It may be that age-related deficiency in a central processing factor set restrictions on the ability to encode/retrieve the type of information that evoke remember experiences. Age-related reductions in processing capacity may, in line with a production deficiency hypothesis, exert a negative influence on performance selectively in the non-guided condition. Thus, the existence of a processing deficiency does not, as such, preclude the existence of a production deficiency. However, it is also possible that age-related reductions in remembering cannot be compensated for by provision of support in the form of encoding instructions and materials; and that the age differences in remembering are due to a global processing deficiency.

The potential influence of age-related reductions in executive/prefrontal functions has been investigated in two prior studies. Parkin and Walter (1992) found that some measures of executive functioning were associated with the extent to which elderly rely on remembering, whereas other measures were not. In a similar vein, Perfect and Dasgupta (1997) found that frontal lobe measures predicted the age-related variance in 2 out of 18 cases. In light of the mixed success of predicting age-related differences in remembering from these measures, we wished to examine the influence of another general factor, processing speed. The processing speed theory posits slowing of elementary processing operation as a major determinant of age-related differences across a variety of cognitive tasks (Birren & Fischer, 1995; Salthouse, 1996) including measures of executive function (Salthouse, Fristoe, & Rhee, 1996; Salt house et al., 2000). As regards episodic memory performance, several studies also suggest that a considerable amount, but not all, of the age-related variance is shared with individual differences in speed measures (e.g., Hultsch, Hertzog, & Dixon, 1990; Lindenberger, Mayr, & Kliegl, 1993; Salthouse, 1993). In line with these prior results, processing speed measures might also serve as an important determinant of age-related differences in recollective experience.

To summarize, we approached the issue of age differences in recollective experience by examining the influence of experimental as well as individual-difference factors. This approach was taken in order to evaluate the relative merits of a production and a processing deficiency account of the age-related decrease in remembering. In a first step we examined the overall pattern of age differences in remembering and knowing, and compared these patterns following a guided (enactment) and a non-guided encoding condition. In a second step, we examined the amount of variance shared by age and processing speed in predicting remembering. We predicted that the results would reveal
age-differences following enacted and non-enacted encoding, respectively. From a production deficiency hypothesis it would be expected that age differences should be eliminated, or substantially attenuated, in the enacted condition. This pattern should be especially clear when remember judgements are considered. Furthermore, we expected that age and processing speed would share a considerable amount of variance in predicting remembering.

Finally, a few words about measurement issues might be in order. Specifically, two major measurement approaches to the first person reports of recollective experience have been advocated in the literature (Donaldson, 1996; Gardiner, Ramponi, & Richardson-Klavehn, 1998; Gardiner, Richardson-Klavehn, & Ramponi, 1998; Hirshman, 1998; Hirshman & Master, 1997). A majority of the prior studies have used raw scores (i.e., number of R-reports) as the dependent measure, sometimes correcting for response bias with a high-threshold model. Others (e.g., Donaldson, 1996) have applied a signal detection model with an accompanying discrimination index (e.g., A'). It is beyond the scope of this article to discuss the relative merits of these approaches, but by reporting both measures when suitable we incorporate both of them. Because the measurement status of the K-reports is less clear (e.g., Gardiner, Ramponi, et al., 1998; Jacoby et al., 1997), we report only the raw scores (cf., Perfect & Dasgupta, 1997).

METHOD

Participants
The participants in the present study were from a sub-sample of the Betula study (Nilsson et al., 1997). The sample originally consisted of 600 individuals from 12 different age cohorts (35, 40, 45, . . . , 90 years). These participants were randomly drawn from the population registry in Umeå, a city of about 100,000 inhabitants in Northern Sweden. The participants were all screened for dementia, severe sensory handicaps, and mental retardation during the recruitment phase (see Nilsson et al., 1997 for further details concerning recruitment and exclusion criteria used). Within this sample, 338 individuals were tested with the R/K procedure. Out of these, 15 were excluded from the study because of failure of understanding of the test instructions. For the statistical analyses, the 12 age cohorts formed six age groups; 35–40, 45–50, 55–60, 65–70, 75–80, 85–90 year-olds. Table 2 provides information concerning participant characteristics and relevant background variables across the studied age groups. As expected from the characteristics of the target population, and as discernable from the means for the present sample, level of educational attainment (self-reported number of years in formal education) was negatively related to age, \( F(5, 317) = 66.80, MSE = 8.76, p < .01 \).

Materials
A total of 64 nouns, belonging to eight different semantic categories served as the to-be-remembered materials. The nouns were divided into four lists of 16 items each. Two lists served as target materials during the study phase, and two lists composed the distractor materials during test. The nouns in each of the lists were dividable into four semantic categories (animals, kitchen utensils, reading materials, and articles of clothing in one target and one distractor list, and parts of the body, musical instruments, carpentry tools and fruits, in the other two). The degree of typicality did not vary across lists (Nilsson, 1973).

In the target lists, each noun was paired with a verb to form unique sentences in imperative form (e.g., lift the book). For each list, eight random orders of items were prepared. The two groups of target lists were counterbalanced across encoding conditions. Objects

Table 2. Participant Characteristics and Background Variables across Age Groups.

<table>
<thead>
<tr>
<th>Age group</th>
<th>n</th>
<th>m/f</th>
<th>Education*</th>
<th>BD-speedb</th>
<th>Letter-digitc</th>
</tr>
</thead>
<tbody>
<tr>
<td>35–40</td>
<td>87</td>
<td>42/45</td>
<td>14.1 (3.2)</td>
<td>8.1 (4.9)</td>
<td>36.4 (6.3)</td>
</tr>
<tr>
<td>45–50</td>
<td>91</td>
<td>47/44</td>
<td>13.6 (3.0)</td>
<td>9.5 (4.5)</td>
<td>33.4 (6.1)</td>
</tr>
<tr>
<td>55–60</td>
<td>58</td>
<td>26/32</td>
<td>11.1 (3.6)</td>
<td>12.3 (7.2)</td>
<td>30.7 (5.7)</td>
</tr>
<tr>
<td>65–70</td>
<td>24</td>
<td>16/8</td>
<td>7.5 (1.7)</td>
<td>20.0 (11.0)</td>
<td>23.3 (6.8)</td>
</tr>
<tr>
<td>75–80</td>
<td>28</td>
<td>15/13</td>
<td>6.5 (2.6)</td>
<td>32.0 (18.3)</td>
<td>19.1 (6.4)</td>
</tr>
<tr>
<td>85–90</td>
<td>35</td>
<td>17/18</td>
<td>6.6 (2.0)</td>
<td>38.4 (18.5)</td>
<td>12.7 (5.7)</td>
</tr>
</tbody>
</table>

*Note. SDs in parenthesis. *Number of years in formal education. bSeconds to completion. cNumber of correct pairings.
corresponding to nouns were used for the enactment condition and the imperatives to be read were presented in black capital letters printed on white cards.

Four test lists, each containing 32 nouns were constructed. The lists were composed of 8 items from each of the four lists described above (i.e., from the two target and the two distractor lists). Specifically, the studied target lists (each containing 16 items) were split into two sets. These sets of 8 target items were counterbalanced across encoding conditions such that each set of items served as targets in the recognition test an equal number of times. The distractors were assigned to each type of encoding according to their category belongings and according to which type of encoding this category derived from.

**Procedure**

The tasks included in the present study were a part of a larger battery of tests, assessing various cognitive functions (see Nilsson, 1999; Nilsson et al., 1997 for a full description). The study-phase was introduced approximately 5 min into the test-session. Each participant was presented with two consecutive lists of to-be-remembered imperatives presented at the rate of 8 s/item. During study of one of the lists, the participant was requested to perform the action described by the imperatives, using the external object corresponding to the noun. The imperatives were read aloud by the experimenter, and the objects were only visible during presentation. No specific instruction of how to encode the imperatives was given for the other list (nondenacted encoding), which was presented visually (on cards) and orally (read aloud by the experimenter). Order of encoding conditions was counterbalanced across participants. Following presentation of each study list an immediate free recall test was given, followed by a cued recall test of the nouns from both lists, data of which are not reported further here.

The test-phase was introduced following a retention interval of approximately 25 min. During this interval, the participants underwent various cognitive tests. At test, the nouns were presented visually (on cards) and were read aloud by the experimenter. The participant was instructed to respond with a “yes” if she/he recognized the noun as one studied previously, and to respond with a “no” if not. The response deadline was set to 5 s/item. Having completed the test, the R/K procedure was introduced, with instructions partly modeled after Gardiner (1988) and Mäntylä (1993). The participant was told that there are two types of recognition experiences, and that remembering meant that the recognition of an item evoked some specific recollection from the learning phase, such as an image, association or thought. An experience of knowing meant that no such experience occurred. Instead the recognition of an item was accompanied by a feeling of familiarity. In addition, the participant was instructed to report if she/he had an unclear experience. That is, in case an item neither evoked an experience of remembering nor knowing. The reason for including this response category was to provide an option for the participants, when their original recognition decision was based on guessing (cf., Mäntylä, 1993). Next, the experimenter read each of the previously recognized nouns aloud and the participant was required to make a R/K/U judgement for each of these.

**Speed Measures**

Two measures were used as indices of processing speed. The first was the Letter Digit Substitution Test, which is a modified version of the Digit Symbol Substitution Test (Salthouse, 1992; Wechsler, 1981). This test comprises a reference table of nine letter-digit pairings. Below these, the letters appear randomized in rows with blank boxes beneath. Participants are required to write down the digits in the empty boxes, according to the reference table. Following a brief practice trial, participants were allotted 60 s to complete as many letter-digit pairings as possible. The number of correct pairings served as the dependent measure.

The second index of processing speed was derived from the Block Design test. It was reasoned that the time to complete the simplest of the timed designs (design three), which is passed by the absolute majority of participants in all age groups (Wechsler, 1981), may serve as an index of processing speed. The correlation between the two speed measures \( r = .68, p < .01 \) suggest that the measures reflect a common factor. For the few participants \( n = 15 \) who did not manage to solve the design within the time limit, the time was set to maximum (60 s).

**RESULTS**

First we present the results of the analyses for overall recognition data, then we turn to analyses of these data as decomposed into R- and K-reports. Finally, commonality analyses including age, education, and speed as predictors of remembering are presented. Only results below alpha-level .01 are reported throughout the ana-

---

1 In fact, a Swedish word resembling the English word “familiar” was used to categorize experiences of knowing. This was done in order to avoid the risk of confounding type of recollective experience and confidence. Specifically, the Swedish word for know (vet) has a connotative ring of certainty to it.
Table 3. Recognition Performance across Age and Type of Encoding (M, SD).

<table>
<thead>
<tr>
<th>Age group</th>
<th>Hits</th>
<th></th>
<th></th>
<th></th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enacted</td>
<td>Nonenacted</td>
<td>Enacted</td>
<td>Nonenacted</td>
<td>Enacted</td>
</tr>
<tr>
<td>35–40</td>
<td>.95 (.09)</td>
<td>.83 (.18)</td>
<td>.05 (.08)</td>
<td>.12 (.13)</td>
<td>.94 (.04)</td>
</tr>
<tr>
<td>45–50</td>
<td>.96 (.07)</td>
<td>.79 (.18)</td>
<td>.06 (.10)</td>
<td>.14 (.17)</td>
<td>.94 (.04)</td>
</tr>
<tr>
<td>55–60</td>
<td>.95 (.10)</td>
<td>.73 (.22)</td>
<td>.10 (.12)</td>
<td>.16 (.17)</td>
<td>.93 (.04)</td>
</tr>
<tr>
<td>65–70</td>
<td>.93 (.12)</td>
<td>.61 (.21)</td>
<td>.13 (.12)</td>
<td>.11 (.14)</td>
<td>.92 (.04)</td>
</tr>
<tr>
<td>75–80</td>
<td>.88 (.13)</td>
<td>.58 (.25)</td>
<td>.14 (.11)</td>
<td>.13 (.21)</td>
<td>.89 (.06)</td>
</tr>
<tr>
<td>85–90</td>
<td>.86 (.15)</td>
<td>.42 (.28)</td>
<td>.21 (.22)</td>
<td>.14 (.18)</td>
<td>.85 (.12)</td>
</tr>
</tbody>
</table>

Note. SDs in parentheses. FA = False Alarms.

yses, due to the large sample size. Preliminary analyses revealed no effects of sex. Effects of study order among conditions (i.e., enactment, nonenactment) were negligible. Hence, the data were collapsed across these factors.

Means and standard deviations for overall recognition (hit rate, false alarm rate, and A′) as a function of age and encoding condition are presented in Table 3. In order to examine overall recognition performance, we conducted a 6×2 (Age Group × Type of Encoding [enacted vs. nonenacted]) mixed ANOVA with A′ as the index of discrimination (Snodgrass & Corwin, 1988). Type of encoding was the within-subject factor. This analysis revealed a significant main effect of age, F(5, 317) = 27.13, MSE = .01, η^2 = .30, as well as of encoding type, F(1, 317) = 241.74, MSE = .00, η^2 = .43. Enacted encoding was associated with a superior level of recognition, and post hoc tests according to the Scheffé method showed that the 85–90-year-olds performed at a lower level than the groups involving 35–70-year-olds. In addition, the 75–80-year-olds performed worse than the groups involving 35–50-year-olds. The interaction between age and type of encoding also achieved significance, F(5, 317) = 4.80, MSE = .00, η^2 = .07. This interaction reflects somewhat less pronounced age effects following enacted encoding. Analyses of simple effects, however, revealed significant age-related differences following enacted, F(5, 317) = 16.34, MSE = .00, η^2 = .20, as well as nonenacted encoding, F(5, 317) = 19.72, MSE = .01, η^2 = .24. Post hoc tests according to Scheffé showed the following results in the enacted condition: 85–90 < 35–40, 45–50, 55–60, 65–70, and 75–80 < 45–50. For the none enacted condition: 85–90 < 35–40, 45–50, 55–60, 65–70, and 75–80 < 35–40. Also, an examination of the individual performance levels suggests that the age by encoding interaction might reflect a ceiling effect for the younger age groups in the enacted encoding condition.

In order to examine this possibility further, we conducted a median-split analysis (e.g., Erngrund, Mäntylä, & Rönnlund, 1996). The rationale underlying this analysis is that the restriction in the range of responses present in the overall analysis may be circumvented by considering the below median-split data for the condition approaching ceiling levels. The participants in each age group were consequently divided into two groups: those above and below the median of their age group, respectively, with regard to performance in the enacted condition (observations at the median were randomly distributed to achieve equal number of cases in each subgroup). In the below-median group, the age by encoding interaction fell short of significance, F(5, 155) = 2.19, MSE = .01. This result warrants a cautious interpretation of the interaction present in the analysis of the overall data, which may reflect a ceiling effect in the enacted condition.

2When the hit rate exceeds or equals the FA rate, A′ = .5 + ((Hit–FA) (1+Hit–FA))/4(4Hit (1–FA)). When computing A′ the hit rate equaled (#Hit+.5)/(#Targets+.1) and the FA rate equaled (#FA+.5)/(#Distractors+.1).
Table 4. Remember Judgements across Age and Type of Encoding (M; SD).

<table>
<thead>
<tr>
<th>Age group</th>
<th>Hits</th>
<th>Remembering</th>
<th>A'</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enacted</td>
<td>Nonenacted</td>
<td>Enacted</td>
</tr>
<tr>
<td>35–40</td>
<td>.91 (.16)</td>
<td>.62 (.24)</td>
<td>.02 (.05)</td>
</tr>
<tr>
<td>45–50</td>
<td>.92 (.12)</td>
<td>.53 (.24)</td>
<td>.03 (.06)</td>
</tr>
<tr>
<td>55–60</td>
<td>.89 (.16)</td>
<td>.49 (.26)</td>
<td>.05 (.08)</td>
</tr>
<tr>
<td>65–70</td>
<td>.89 (.16)</td>
<td>.38 (.23)</td>
<td>.06 (.10)</td>
</tr>
<tr>
<td>75–80</td>
<td>.81 (.19)</td>
<td>.38 (.26)</td>
<td>.06 (.08)</td>
</tr>
<tr>
<td>85–90</td>
<td>.70 (.28)</td>
<td>.26 (.27)</td>
<td>.09 (.11)</td>
</tr>
</tbody>
</table>

Note. SDs in parentheses. FA = False Alarms.

Phenomenology of Recognition

The following analyses were based on the absolute proportions, that is, #hits/#targets and #false alarms/#distractors were computed separately for R- and K-reports (e.g., if 4 out of 8 targets were recognized, and 3 out of these were accompanied by R-reports, the proportion (R) is 3/8 = .375). However, to allow for comparison with other approaches to the measures derived from the R/K procedure, a nonparametric discrimination index (A'; e.g., Donaldson, 1996) is reported where appropriate. Means and standard deviations across age groups and encoding types are displayed for the R/K-reports in Tables 4 and 5, respectively.

Each of the dependent measures were subjected to a 6×2 (Age Group×Type of Encoding [enacted vs. nonenacted]) mixed ANOVA, with encoding type as the within-subject factor. The response rate for the unclear category was too low to allow for meaningful analysis (below .05 in all age groups).

Table 5. Know Judgements Across Age and Type of Encoding (M; SD).

<table>
<thead>
<tr>
<th>Age group</th>
<th>Hits</th>
<th>Knowing</th>
<th>FA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enacted</td>
<td>Nonenacted</td>
<td>Enacted</td>
</tr>
<tr>
<td>35–40</td>
<td>.04 (.11)</td>
<td>.19 (.18)</td>
<td>.02 (.05)</td>
</tr>
<tr>
<td>45–50</td>
<td>.04 (.08)</td>
<td>.25 (.19)</td>
<td>.03 (.07)</td>
</tr>
<tr>
<td>55–60</td>
<td>.05 (.10)</td>
<td>.22 (.18)</td>
<td>.03 (.07)</td>
</tr>
<tr>
<td>65–70</td>
<td>.03 (.08)</td>
<td>.18 (.14)</td>
<td>.04 (.07)</td>
</tr>
<tr>
<td>75–80</td>
<td>.07 (.08)</td>
<td>.15 (.14)</td>
<td>.04 (.08)</td>
</tr>
<tr>
<td>85–90</td>
<td>.15 (.25)</td>
<td>.14 (.16)</td>
<td>.09 (.17)</td>
</tr>
</tbody>
</table>

Note. SDs in parentheses. FA = False Alarms.
ing. Analyses of simple effects, however, revealed significant age-related differences following enacted, \(F(5, 317) = 10.50, \text{MSE} = .03, \eta^2 = .14\), as well as nonenacted encoding, \(F(5, 317) = 13.29, \text{MSE} = .06, \eta^2 = .17\). For enacted encoding, the Scheffé test revealed the following group differences: 85–90 < 35–40, 45–50, 55–60, 65–70. The corresponding analyses for nonenacted encoding showed that: 85–90 < 35–40, 45–50, 55–60, and 65–70, 75–80 < 35–40. Analyses with the A’ discrimination index as the dependent measure revealed no deviations from the pattern reported above.

In order to examine the validity of our suspicion that the age by encoding interaction reflected a ceiling effect for the younger adults in the enacted condition, we once again conducted a median-split analysis. The participants in each age group were divided into two groups based on correct remembering in the enactment condition; those above the median of their age group and those below the median. The outcome of this split for the below-median group is depicted in Figure 1. The means are suggestive of parallel age-related differences across encoding conditions, and, consistent with this observation, the age by encoding interaction fell short of significance, \(F(5, 155) = 2.15, \text{MSE} = .03\). Thus, the age by encoding interaction for correct remembering present in the overall analysis should be interpreted with caution, and may well reflect a ceiling effect for the enacted condition.

The analysis of false remembering (i.e., false alarms) revealed main effects of age, \(F(5, 317) = 5.11, \text{MSE} = .01, \eta^2 = .08\), and encoding, \(F(1, 317) = 7.74, \text{MSE} = .00, \eta^2 = .02\). These effects come as the result of an age-related increase in false remembering and a higher rate of false remembering for distractors in the enacted condition, as seen in Table 4. However, the Scheffé test showed no age differences between the groups. The interaction between age and encoding was significant, \(F(5, 317) = 3.28, \text{MSE} = .00, \eta^2 = .05\). Tests of simple effects yielded a significant increase with higher age for distractors in the enacted condition, \(F(5, 317) = 6.61, \text{MSE} = .01, \eta^2 = .09\). By contrast, no such differences were found for distractors in the nonenacted condition. The Scheffé test revealed that the 85–90-year-olds reported more false remembering following enacted encoding than the 35–40- and 45–50-year-olds.

**Knowing**

An inspection of the means for correct knowing (i.e., hit rate) summarized in Table 5 suggests that participants had fewer know experiences for enacted items. The ANOVA results confirmed this impression, yielding a main effect of encoding, \(F(1, 317) = 103.17, \text{MSE} = .02, \eta^2 = .25\). The main effect of age was not significant, \(F<1\). However, the age by encoding interaction reached significance, \(F(5, 317) = 7.25, \text{MSE} = .02, \eta^2 = .10\). Simple effects analyses revealed an increase of correct knowing with advancing age following enacted encoding, \(F(5, 317) = 4.91, \text{MSE} = .02, \eta^2 = .07\), and a decrease with higher age following nonenacted encoding, \(F(5, 317) = 3.13, \text{MSE} = .03, \eta^2 = .05\). For enacted encoding, the Scheffé test showed that the 85–90-year-olds reported more correct knowing than the 35–40- and 45–50-year-olds. The corresponding test for nonenacted encoding revealed no group differences. The analyses of false knowing yielded a main effect of encoding only, \(F(1, 317) = 19.28, \text{MSE} = .01, \eta^2 = .06\). This effect demonstrated that the participants gave fewer false know judgements for items belonging to the categories encoded with enactment.

In addition to these analyses we conducted planned comparisons, comparing correct and
false knowing within both encoding conditions. The rationale for these analyses was that it would only be justified to draw any firm conclusions regarding knowing if the hit rate exceeded the false alarm rate. The analyses revealed that the levels of correct knowing did not differ from the levels of false knowing following enacted encoding. Thus, the participants did not report levels of knowing above chance performance. By contrast, following non enacted encoding, correct knowing was higher then false knowing, \( t(322) = 10.90 \).

**Commonality Analyses**

First, we averaged the \( z \) scores of the speed measures to form a composite measure and collapsed remembering across encoding conditions. Due to the well-established association between education and cognition (e.g., Inouye, Albert, Mohs, Sun, & Berkman, 1993) and the marked differences in educational attainment across age groups within the present sample, we included years of schooling in the analyses. Missing values for education (\( n = 11 \)) were replaced with the sample mean for the highest educational level attained by the participant (e.g., high school grade), or, in case this information was missing, the mean of the age cohort to which he/she belonged. Zero-order and age-partialled correlations among the variables are presented in Table 6. Notably, correlations among the various memory measures were substantial. Age was negatively associated with processing speed, and there was a positive association between each of the memory measures and speed, also when age was partialled out. Quadratic trends in the variables were examined by entering the squared age term, after the linear age term in the regression equations. Because the quadratic trends were small or nonexistent (\( R^2 < .027 \)), they were ignored.

To decompose the total variance accounted for by the three hypothesized predictors (age, processing speed, and education) in measures of remembering (hits) and recognition (A'), we regressed the respective criterion on all seven possible predictor combinations. Of primary interest, age explained about 20.9% of the variance in remembering and approximately 25.9% of the variance in recognition. The total amount of variance accounted for by age, education, and processing speed, together, was 25.2% for remembering and 30.1% for recognition. The amount of variance accounted for by each of the seven equations was then used to compute unique and shared variance components predicting individual differences in remembering and recognition, respectively (cf., Lindenberger & Baltes, 1994; Pedhazur, 1982). The results are presented in Table 7. It should be noted that use of A' as the measure of remembering showed results very similar to that reported here, as did separate analyses for the enacted and non enacted conditions.

The three predictors shared approximately 56% of the explained variance in remembering and about 58% of variance in recognition. In addition, age and processing speed shared about 21% of the explained variance in remembering and 17% in recognition, beyond the variance shared with education. All other variance components were

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>–</td>
<td>– .70**</td>
<td>– .82**</td>
<td>– .53**</td>
<td>– .51**</td>
<td>– .46**</td>
<td>– .47**</td>
</tr>
<tr>
<td>2. Edu.</td>
<td>–</td>
<td>–</td>
<td>– .48**</td>
<td>– .48**</td>
<td>– .41**</td>
<td>– .42**</td>
<td></td>
</tr>
<tr>
<td>4. Rn (Hits)</td>
<td>–</td>
<td>.18***</td>
<td>– .23**</td>
<td>–</td>
<td>.73**</td>
<td>.78**</td>
<td>.71**</td>
</tr>
<tr>
<td>5. Rn (A')</td>
<td>–</td>
<td>.20***</td>
<td>– .16**</td>
<td>– .63**</td>
<td>– .70**</td>
<td>.77**</td>
<td></td>
</tr>
<tr>
<td>6. R (Hits)</td>
<td>–</td>
<td>.14***</td>
<td>– .20**</td>
<td>– .71**</td>
<td>– .61**</td>
<td>–</td>
<td>.91**</td>
</tr>
<tr>
<td>7. R (A')</td>
<td>–</td>
<td>.15***</td>
<td>– .20**</td>
<td>– .62**</td>
<td>– .70**</td>
<td>– .88**</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note. Recognition and remembering are collapsed across type of encoding. Speed is a \( z \) score composite. Rn = recognition; R = remembering; Edu. = years in formal education.

*p < .05. **p < .01.
Table 7. Commonality Analyses: Shared and Unique Components in Predicting Recognition and Remembering.

<table>
<thead>
<tr>
<th>Component</th>
<th>Variance explained (%)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recognition (A')</td>
<td>Remembering (hits)</td>
</tr>
<tr>
<td>Unique age</td>
<td>1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Unique education</td>
<td>2.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Unique speed</td>
<td>1.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Shared age, speed</td>
<td>5.0</td>
<td>5.3</td>
</tr>
<tr>
<td>Shared age, education</td>
<td>2.3</td>
<td>1.0</td>
</tr>
<tr>
<td>Shared education, speed</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Shared age, education, speed</td>
<td>17.5</td>
<td>14.2</td>
</tr>
<tr>
<td>Total variance</td>
<td>30.1</td>
<td>25.2</td>
</tr>
</tbody>
</table>

relatively small. Taken together, the commonality analyses reveal that processing speed and age overlap to a considerable extent in the prediction of remembering and recognition. Specifically, 77% (56 + 21) of the explained variance in remembering and 75% (58 + 17) in recognition is shared among age and speed. Furthermore, 93% of the age-related variance in remembering and 87% of the age-related variance in recognition were shared with processing speed.

DISCUSSION

Several clear-cut and potentially important results were obtained. These concern (a) the age-related patterns of recollective experience, (b) the effect of enactment on these experiences and upon the age differences, and (c) the relation between age-related differences in remembering and individual differences in basic processing efficiency, as indexed by measures of processing speed. We address these issues in turn.

An age-related reduction in the probability that an item will evoke a state of conscious recollection, that is being remembered, was observed. This result is in line with the more general trend in prior studies (see Table 1). Our findings extend the generality of past findings by demonstrating a continuous age-related decrement in R-report levels from 35 to 90 years of age. The incidence of false remembering (i.e., remember reports for distractors) tended to increase with advancing age, a finding that is in concordance with prior results (e.g., Norman & Schacter, 1997; Schacter et al., 1997). The magnitude of age differences in remembering did not deviate from that observed in overall recognition performance. Thus, one general point raised by the present results is that attempts at "purifying" recognition measures in terms of episodic content by exclusively considering remember reports, do not basically alter the magnitude of the age effects in recognition. In other words, the results suggest that aging is associated with quantitative, rather than qualitative, changes in recognition.

Turning to the effects of the encoding manipulation, the results confirmed previous findings of an impressive enactment effect for R-reports (Conway & Dewhurst, 1995; Engelkamp & Dehn, 1997). In fact, the magnitude of this effect was larger in terms of R-reports than in overall recognition performance. Recent theoretical developments emphasize distinctive processing as a key construct to explain such encoding effects (Rajaram, 1996, 1998). In a similar vein, the enactment effect has typically been attributed to increased processing of item-specific information (e.g., Engelkamp & Zimmer, 1984) among which motor information appears to be critical (Heil et al., 1999). Tactile cues and richness of sensory features associated with manipulation of concrete objects (cf., Bäckman & Nilsson, 1984; Engelkamp & Zimmer, 1997), probably served to enrich the studied events further. Of primary concern within the present context, the results show that our guided encoding manipulation had a powerful effect on subsequent memory performance, especially as reflected by remember experiences.

As outlined in the introduction, the modifiability of age differences in episodic memory by encoding factors is abstruse, in light of evidence from overall performance studies (e.g., Burke & Light, 1981; Craik & Jennings, 1992; Light, 1991), as well as in light of previous R/K studies (see Table 1). The present findings discourage the view that the magnitude of age differences in recollective experience is determined by whether encoding is guided or not (Perfect et al., 1995). This is because age-related deficits in levels of
R-reports were clearly present following non-enacted as well as enacted encoding. Thus, whereas prior studies (Table 1) suggested that age-related differences were eliminated, or attenuated substantially, following guided encoding, this study reveals highly significant age differences following guided as well as non-guided encoding. Attempts to remedy a possible ceiling effect in the enacted condition by a median-split analysis in fact suggest that the magnitude of the age effects is independent of encoding type. The fact that the median-split sample in the present study is larger than that included in all comparable R/K-studies suggest that the decreased power associated with this procedure might be of minor importance. In addition, previous studies generally confirm that the lower performing half of the population is not selectively responsible for age-related differences in cognitive performance (see Salthouse, 1991 for review).

The fact that our results, overall, indicate somewhat larger age-related deficits than most prior studies ($M = .16$) is possibly due to inclusion of old-old participants (> 80 year old). The fact that the present results were obtained for population-based rather than education-matched age samples, may be another factor. The finding of a similar age-related pattern with regard to recognition performance and remembering following enacted and nonenacted encoding converges with the majority of studies of recall performance (Rönnlund et al., 2002). Thus, whereas enactment constitutes a very effective form of memory support, it appears to be utilized to about the same extent regardless of age.

The incidence of false remembering (i.e., remember reports for distractors) tended to increase with advanced age (cf. Norman & Schacter, 1997; Schacter et al., 1997), an effect that was restricted to the enacted condition. As the items were organiziable along semantic categories and the distractors were from the same categories, this finding possibly signals intact semantic categorization in conjunction with inability of making the more fine grained within-category discriminations necessary to distinguish between presented/nonpresented items (cf., Koutstaal & Schacter, 1997). The fact that this effect occurred only for the enacted items is intriguing, but difficult to disentangle within the scope of the present article.

Turning to know reports, an age-related deficit in hit rate was observed following nonenacted encoding, but the magnitude of this effect was weak. Because the measurement status of the K-reports is currently somewhat unclear, the theoretical significance of this result is difficult to elaborate further. The tendency for increased levels of knowing following enactment scarcely captures theoretical interest, since correct knowing did not exceed the false alarm rate in this particular case. The fact that remember levels approached ceiling levels for the young in this condition, leaving little room left for knowing, furthermore precludes interpretation of the data. Even if we restrict the analysis to the nonenacted condition, however, the present results do not confirm previous accounts of the differential age-related patterns across studies. That is, because knowing increases following maintenance rehearsal (Gardiner et al., 1994), an age-related increase would have been expected if older adults are more likely to engage in maintenance rehearsal (Perfect & Dasgupta, 1997; Perfect et al., 1995). Hence, the present finding, that know-reports are relatively age-insensitive (or even decrease), is inconsistent with a production deficiency interpretation.

The impression that age differences in recollective experience are determined by the nature of the encoding task possibly emanates from procedural differences across studies. This argument is endorsed by recent evidence of the sensitivity of the R/K procedure to different testing procedures (Hicks & Marsh, 1999). A few aspects in which the procedure of the present study differed from that used in previous ones may therefore be noteworthy.

The recognition test required words (nouns) to be recognized, as in past studies, but these were encoded in the context of a verb-noun sentence. It is possible that this affected the strategy adopted at time of encoding to some extent. Specifically, participants may have been less likely to rely on a maintenance rehearsal strategy in the non-guided condition in the present case. As a result, the contrast between the guided and the non-guided conditions may have been less than maximal by virtue of inherent task properties. Still, one would
expect a maintenance rehearsal strategy to be more prevalent in the nonenacted than in the enacted condition, even if one does not consider the nonenacted condition to be "purely" non-guided. Furthermore, the effects of the presumably more important variation, that is the presence or absence of specific encoding instructions which has been used to define the guided/non-guided distinction in previous R/K-studies (Perfect et al., 1995), needs to be accounted for. If valid, the production deficiency hypothesis would predict an age by encoding interaction rather than lack of one, as observed at present. One may, in addition, note that the recognition test was preceded by tests of recall of the same materials, a procedure that deviates from prior studies. We acknowledge the possibility that this may have served to magnify overall R-levels. Specifically, prior recall may serve as an additional opportunity of encoding aspects of the original events, information that may subsequently trigger remembering. On the basis of findings of equal effects of prior retrieval in young and old adults, both in recall of verbal materials (Rabinowitz & Craik, 1986) and enacted sentences (Kausler & Wiley, 1991), we do not believe that this fact qualifies our conclusions with regard to the magnitude of age-related differences observed. Rather, one may speculate that scaling effects, specifically those associated with near ceiling performance in some guided conditions, may be a factor to consider when interpreting the patterns observed across prior studies. The fact that we, unlike previous studies, base our conclusions on within-subjects data from a representative sample that covers a wide age range should ascertain that the present findings are fairly general.

The results of the present study may appear pessimistic as regards the possibility of removing age effects by use of simple interventions, such as specific encoding instructions. However, it should be pointed out that our results demonstrate considerable plasticity in old age; following provision of support older adults are at levels with, or even outperform, the young in the non-guided condition, both in terms of overall recognition and remember experiences. Also, the result of highly similar age differences for the guided and non-guided encoding is not incompatible with the notion that deficits in recollective experience in old age emanate from encoding problems. That is, older adults may encode information less efficiently relative to younger adults, regardless of whether encoding is guided or not. As discussed above, there appear to be several results consistent with this position. Whereas the present study did not address the issue of the locus of the age differences directly, it may be noted that the magnitude of the age effects in recognition observed at present are not too different in magnitude from those observed for measures of recall (Rönnlund et al., 2002). Thus, reducing the demands for retrieval by requesting recognition rather than recall, only marginally modifies the magnitude of age differences, a finding which suggests that differences in encoding processes are central for understanding age-related memory loss.

If presence or absence of specific instructions is not an important determinant of age-related differences in memory, what is? In the light of the weak support for a production deficiency interpretation of the age-related deficits in remembering we examined the extent to which the age-related variance in remembering was shared with measures of processing speed. In line with several prior studies (e.g., Hultsch et al., 1990; Lindenberger et al., 1993; Salthouse, 1993), we found that individual differences in speed and age-

\[\text{Contingency analyses involving recognition data and data from the second (cued-) recall test, which, like the recognition test, involved the noun (the prior free recall test recall of the whole sentences was requested), revealed that the proportion R-hits were higher for previously recalled than for previously nonrecalled items, across age and encoding type, possibly related to the fact that prior retrieval increased the opportunities to encode information that subsequently triggered remembering. Negative age differences in R-hits were present both for previously recalled and nonrecalled items, however, suggesting that potential effects of prior retrieval generalized across encoding type. It should be noted that these results are restricted by the fact that they are correlational in nature; higher frequency of R-reports for previously recalled items could alternatively reflect the fact that these were differently encoded in the first place, such that they were easier to recall and more likely to be remembered in the subsequent recognition test as compared with the nonrecalled items.}\]
related variance in memory performance overlapped considerably. The finding that most of the age-related variance in memory is shared with measures of processing speed is attributable to several different accounts. For example, speed might be a sensitive indicator of a yet unspecified common factor, such as brain integrity (cf. Lindenberger & Baltes, 1994).

It could also be argued that the association between processing speed and memory was inflated by the use of a timed stimulus presentation/response format. In general, however, task pacing does not stand out as a critical determinant of age differences in episodic memory (Verhaeghen, Marcoen, & Goosens, 1993). If anything, some studies show a tendency such that slower presentation magnifies the age differences (e.g., Craik & Rabinowitz, 1985; Hartley, Stojack, Mushaney, Annon, & Lee, 1994), a finding which may be hypothesized to reflect younger adult’s superior ability to engage in further elaborative or distinctive processing. That is, processing speed may set restrictions not only as to the rate of rehearsal, but also as to the breath, or extensiveness of encoding that will be achieved (Salthouse, 1994). Hence, additional time for encoding may sometimes be utilized more efficiently by younger adults.

Furthermore, there are quite a number of studies suggesting that age-related cognitive slowing exerts at least part of its influence on episodic memory performance indirectly via reductions in other capacity delimiting factors such as working memory and inhibition (e.g., Park et al., 1996; Van der Linden et al., 1999). Hence, the present result is not inconsistent with the view that age-related deficits in recollective experience emanate from impairments in other factors such as working memory and executive function. Future studies, including careful operationalizations of multiple constructs, may shed further light on these issues.

In conclusion, enacted encoding was found to yield superior levels of remembering as compared with a condition in which the equivalent information was presented without encoding support. Age-related deficits in recollective experience were found to generalize across this manipulation, suggesting that guiding the encoding by requesting enactment does not serve to alter the magnitude of the age effects in any substantial fashion. Thus, a production deficiency hypothesis does not seem viable from the point of the present data. Rather, our results suggest that these age differences are better predicted by reduction in some more general factors such as those indicated by measures of processing speed.

ACKNOWLEDGEMENT

This research was supported by the Bank of Sweden Tercentenary Foundation, the Swedish Council for Planning and Coordination of Research, the Swedish Council for Research in the Humanities and Social Sciences, and the Swedish Council for Social Research. We thank Lars Nyberg for comments on earlier drafts of this manuscript.

REFERENCES


Selective Adult Age Differences in an Age-Invariant Multifactor Model of Declarative Memory

Lars Nyberg
Umeå University

Michael Rönnlund
Umeå University

Scott B. Maitland
University of Guelph

Lars Backman
Karolinska Institute

Roger A. Dixon
University of Alberta

Ake Wahlin
Karolinska Institute

Lars-Göran Nilsson
Stockholm University

Confirmatory factor analysis was used to test competing models of declarative memory. Data from middle-aged participants provided support for a model comprised of 2 2nd-order (episodic and semantic memory) and 4 1st-order (recall, recognition, fluency, and knowledge) factors. Extending this model across young-old and old-old participants established support for age invariance. Tests of group differences showed an age deficit in episodic memory that was more pronounced for recall than for recognition. For semantic memory, there was an increase in knowledge from middle to young-old age and thereafter a decrease. Overall, the results support the view that episodic memory is more age sensitive than semantic memory, but they also indicate that aging has differential effects within these 2 forms of memory.

In this article, we address the structure and development of declarative memory in middle-aged and older adults (age range 35–80 years). After establishing structural characteristics of declarative memory, our principal goals included (a) testing for measurement invariance of the structure across three adult age groups and (b) examining age differences in performance at the level of latent factor means.

An enduring debate in the psychology of memory concerns the distinction between semantic and episodic memory. Briefly, the issue is whether memory for general fact information and memory for more specific event information represent two separate memory systems, or whether they are related manifestations of a more general declarative memory system (e.g., Squire & Knowlton, 1995; Tulving, 1983, 1995). With respect to memory and aging, the discussion has concerned (a) whether these two apparent systems may be identified not only with young adults but throughout late life and (b) the extent to which one system (i.e., episodic) may be more aging sensitive than the other (i.e., semantic). These issues have been addressed in cross-sectional comparisons of older and younger adults (e.g., Nilsson et al., 1997), empirical reports of longitudinal investigations (e.g., Dixon et al., in press; Hultsch, Hertzog, Dixon, & Small, 1998), as well as in reviews of particular subgroups such as middle-aged (Dixon, de Frias, & Maitland, 2001) and very old (e.g., Backman, Small, Wahlin, & Larsson, 1999) adults.

Evidence that declarative memory can be divided into two different systems comes from many sources, including analyses of covariance relations between memory measures. Using covariance-based statistical techniques, one testable assumption is that scores on measures of the same underlying factor/system would be
more strongly interrelated than scores on measures of different underlying factors/systems. In an early study, Underwood, Boruch, and Malmin (1978) compared the intercorrelations between scores on episodic and semantic memory tasks. They concluded that “our episodic memory tasks and the semantic memory tasks represent two different worlds” (p. 409). Comparison of the covariance structure associated with episodic and semantic measures can also be done by factor-analytic methods. Using principal-components factor analysis, Mitchell (1989) observed that measures of episodic and semantic memory loaded on separate factors (see also Nilsson et al., 1997). Early work by Thurstone (1943) can also be taken as factor-analytic support for the semantic-episodic distinction. Factor analysis is often exploratory in nature, with no a priori assumptions about the relationship between measures and underlying factors built into the analysis. In confirmatory factor analysis, however, hypotheses about an underlying factor structure and its relation to observed variables can be examined. Nyberg (1994) used confirmatory factor analysis to compare a unitary model of declarative memory with a two-factor episodic-semantic model (episodic memory was measured by recall and recognition tests; semantic memory was measured by tests of general and word knowledge). Using structural equation modeling (SEM; Jöreskog & Sörbom, 1989), support was obtained for the two-factor model. Thus, the results from a variety of correlational studies are consistent with the notion of differentiation between episodic and semantic memory systems.

In the present study, covariance relations between measures of semantic and episodic memory in three adult age groups were examined further. The first goal was to examine whether a two-factor (semantic-episodic) model of declarative memory provided a better fit than a unitary model (cf. Nyberg, 1994). Although the present sample is drawn from the same population as in the Nyberg (1994) study, it constitutes an independent sample covering a broader range of adulthood. The second goal was to examine the support for further division of semantic and episodic memory into subsystems. The possibility that semantic and episodic memory consist of subsystems has long been discussed (e.g., Tulving, 1987). There is empirical support for a division of episodic memory into recall and recognition from functional (Gregg, 1976), brain-damage (Hirst, Johnson, Kim, Phelps, & Volpe, 1986; Hirst, Johnson, Phelps, & Volpe, 1988) and brain-imaging (Cabeza et al., 1997) studies. As concerns semantic memory, an old subdivision from the literature of psychometric intelligence, based on exploratory factor analytic studies, is between verbal comprehension and verbal fluency (Thurstone, 1943). Brain-damage data also indicate that performance in tasks requiring speeded verbal production, such as word fluency, are selectively disrupted following frontal—especially left-sided—lesions, not only in comparison with lesions located elsewhere but also in comparison with other verbal ability measures (e.g., Benton, 1968; Perret, 1974). In a similar vein, effects of closed head injury have been found in tasks involving rapid semantic access, despite intact performance in comprehension tasks, such as vocabulary (Haut, Petros, Frank, & Haut, 1991). Thus, for episodic as well as semantic memory, several lines of evidence are suggestive of dissociable subsystems. However, no study has demonstrated a comprehensive model of declarative memory that provides support for all of these subsystems in any age group, much less in multiple groups of adults of varying ages. Our measurement battery allowed us to test the feasibility of subdividing episodic memory into recall and recognition factors and semantic memory into knowledge and fluency factors.

A third goal of the present research was to examine the validity and generalizability of the final factor solution by testing whether it was age invariant. We tested the final accepted structure by using confirmatory factor analysis with nonzero latent means (Everson, Millsap, & Rodriguez, 1991; Jöreskog & Sörbom, 1979; Millsap & Everson, 1991) on three age groups from the same sample. A model that is well fitting for middle-aged adults may not be representative of the underlying factor structure for older adults. Such an outcome would question the validity of quantitative comparisons of test scores of middle-aged and older adult groups, as it would indicate a differential declarative memory structure across the adult life span. In contrast, if a factor structure demonstrates measurement invariance across age groups, then age-related performance differences can be interpreted as quantitative variations within a common structure.

Given evidence for age invariance, our fourth and final goal was to compare age differences in performance for first- and second-order latent factor means. Of primary interest was whether latent mean differences would demonstrate factor-specific age differences in levels of memory performance. The extant empirical evidence concerning the adult life span trajectories for overall (second-order) episodic and semantic memory functioning indicates that age-related deficits are robust for episodic memory, although they are attenuated or even eliminated for semantic memory (for reviews, see Bäckman et al., 1999; Craik & Jennings, 1992). With regard to the recall-recognition and knowledge-fluency (first-order) subdivisions, the evidence is less clear. Specifically, it is commonly believed that age-related differences are smaller in recognition than in recall (see Spencer & Raz, 1995, for a meta-analysis) because of the greater demands of self-initiated retrieval operations in recall (e.g., Bäckman, Mäntylä, & Herlitz, 1990; Craik & McDowd, 1987). However, at the same time, a clear age-related impairment in face recognition is one of the most consistently observed phenomena in the literature on memory and aging (e.g., Bartlett & Fulton, 1991; Wachtel et al., 1993). In addition, most available evidence suggests equal gains in young and older adults from increasing levels of retrieval support and other environmental guidance (e.g., Bäckman et al., 1999). In an analogous vein, much research demonstrates that age-related differences in the structure of the semantic network are negligible (e.g., Balota & Duchek, 1988; Laver & Burke, 1993), although older adults have problems in accessing lexical information rapidly (e.g., Auci et al., 1995; Maylor, 1990; for a review, see Light, 1992). Yet, some recent work suggests that age-related deficits may be observed in tasks tapping semantic knowledge with limited speed demands, such as vocabulary (e.g., Gilinsky & Judd, 1994; Lindenberger & Baltes, 1994). An important objective of the present study was to formally examine these empirical patterns at the level of latent factor means. Structural equation model development and comparisons proceeded in a predefined sequence of steps. First, the best-fitting model was determined for the middle-aged group. Second, we analyzed the best-fitting model with regard to age-group invariance. Three adult groups were tested (middle-aged, young-old, old-old), controlling for differences in educational attainment among groups. Finally, analyses of latent factor means examined differences in declarative memory performance among the three age groups.
Method

Participants

The participants in the present study were 925 community-dwelling individuals from the Betula project (Nilsson et al., 1997). The ongoing Betula project includes several independent samples. All participants in the present study were from Sample 3 (participants from Sample 1 were included in Nyberg, 1994). Sample 3 was tested for the first time in 1993–1995, and the data reported here are from their first wave of testing. The participants were recruited by random selection of names from the population registry. For the present analyses, the sample was divided into three age groups as follows: middle-aged group, M_age = 40 years (range = 35–50); young-old group, M_age = 60 years (range = 55–65); old-old group, M_age = 75 years (range = 70–80). The proportion of women exceeded that for men in all age groups. Subjective health was assessed by asking participants whether they felt healthy or not (e.g., “yes or no”).

Participants that indicated not healthy increased in older age, whereas level of education decreased. An analysis of variance (ANOVA) showed that the mean years of education were significantly different among all age groups, F(2, 923) = 172.33, p < .001. To assess global level of cognitive functioning, the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) was administered according to standardized procedures. The score used was out of a maximum of 30. An ANOVA showed that the MMSE scores were significantly different among all age groups, F(2, 923) = 57.00, p < .001. Participants with an MMSE score below 24 were excluded (N = 33), as this may indicate dementia (e.g., Small, Viitanen, & Bäckman, 1997). Also, a few individuals had missing observations in the memory tasks and were therefore excluded from the statistical analyses. These exclusions resulted in 377 individuals in the middle-aged group, 283 individuals in the young-old group, and 265 individuals in the old-old group. The sample characteristics are summarized in Table 1.

Measurement Variables

The memory tasks were administered during two test sessions, both of which lasted between 1.5 and 2 hr for each participant.1 The present analyses were based on 12 episodic tests (9 recall tests, 3 recognition tests) and 5 semantic tests (3 fluency tests, 2 knowledge tests). The tests are described below, and means and standard deviations are given in Table 2.

Episodic Tests: Recognition

Face/name recognition. Early in the second test session, participants silently viewed 16 photographs of children’s faces (8 s/photo). Below each photo, a Swedish first and family name was typed. Participants were asked to memorize the face together with the family name. Approximately 45 min into the session (a delay during which other cognitive tests were administered), participants were shown 24 photographs of 12 studied and 12 nonstudied faces, randomly intermixed. The participants were asked to make an old and a new judgment in response to each face. For each face judged to be old and for each missed target face, they were asked to identify the names from a list of four alternatives (first and family name). Number of hits minus false alarms for faces (RN FACES) and number of hits for family names (RN NAMES) were entered in the analyses.

Recognition of nouns. Approximately 60 min into the second test session, participants were presented with a list of 32 nouns. Half of these nouns were from the enacted and nonenacted sentences studied earlier, with 8 nouns from each condition (the 16 other nouns were nonstudied and served as distractors). Number of hits minus false alarms for nouns from the nonenacted condition (RN NOUNS) served as the measure adopted here.

Table 1 Sample Characteristics

<table>
<thead>
<tr>
<th>Age group</th>
<th>Male/ female</th>
<th>Education</th>
<th>MMSE</th>
<th>Healthy/ not healthy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Middle-aged (35–50)</td>
<td>186/191</td>
<td>12.86</td>
<td>3.29</td>
<td>28.36</td>
</tr>
<tr>
<td>Young-old (55–65)</td>
<td>128/155</td>
<td>9.67</td>
<td>3.63</td>
<td>27.94</td>
</tr>
<tr>
<td>Old-old (70–80)</td>
<td>107/158</td>
<td>7.93</td>
<td>3.38</td>
<td>27.14</td>
</tr>
</tbody>
</table>

Note. MMSE = Mini-Mental State Examination (Maximum M = 30).
1 Age is in years. 2 Formal schooling (in years). 3 Proportion of self-reported health.

1 For detailed descriptions of the structure of the test sessions and a complete list of the tests included, consult Nilsson et al. (1997) and Nilsson (1999).
of each list, participants recalled as many of the nouns as possible in any order at a given pace (2 s/item), indicated by a metronome. For one list, the task was performed under conditions of full attention at study and retrieval (RC FF). Study/retrieval of words in the other lists was paired with performing a secondary task. This task consisted of sorting red and black cards into two piles on basis of color (2 s/item). In one condition, division of attention occurred at study of the nouns, but not at retrieval (RC DF). In another condition, participants were requested to sort the cards at retrieval of the nouns, but not at study (RC FD). In a final condition, both study and retrieval of words occurred under conditions of divided attention (RC DD).

Semantic Tests: Knowledge

Vocabulary. A 30-item multiple-choice synonym test (SRB; Dureman, 1960) was used. The task was to select the synonym of each target word from among five alternatives. Seven minutes were allotted for completing the test. The total number of correctly identified synonyms (VOCAB) was entered into the analyses.

General knowledge. The items in this test appeared as part of the test of item/source recall described above. Ten questions included in this test were possible to answer on the basis of acculturated knowledge (e.g., Where was Ingemar Stenmark born?). Number of correctly answered statements within this category (GEN KNOWL) served as the outcome measure.

Semantic Tests: Fluency

Word fluency. Three tests in which the participants generated as many words as possible in 1 min were used. In the first test, the participants said aloud as many words as possible with an initial letter A (FLUA). The second test was to produce as many words as possible beginning with M and containing five letters (FLUM). The third test was to generate as many professions as possible with the initial letter B (FLUB).

Statistical Procedures

Confirmatory factor analysis was conducted with LISREL 8.51 (Jöreskog & Sörbom, 2001). Analyses were conducted on covariance matrices and mean vectors with results of the final models reported as standardized estimates for ease of interpretation. Factor scaling was accomplished by fixing one item for each factor to a value of 1.0 in the pattern matrix (u in LISREL) and the same item was used to scale factors among age groups. The chi-square difference test (H9273\(\Delta \chi^2\); Jöreskog & Sörbom, 1989) was used to compare nested models. The critical value used for all comparisons was \(p < .01\).

Equivalence of factor structures was evaluated for configural and metric invariance (Horn, 1991; Horn, McArdle, & Mason, 1983; Horn & McArdle, 1992; Meredith, 1993). First, configural invariance examined whether the factor patterns were identical across groups (i.e., whether the same variables load on the same number and pattern of factors across age groups). Second, metric invariance requires equivalence of factor loadings across age groups. This was examined sequentially for the first- and second-order factor loadings, respectively. Metric invariance allows for meaningful quantitative comparisons of groups by establishing comparable measurement units for the variables and factors (Cunningham, 1982, 1991; Horn et al., 1983; Horn & McArdle, 1992; Maitland, Dixon, Hultsch, & Hertzog, 2001). Additionally, latent mean structures were examined to determine age-group differences in memory performance. Methods described by Byrne (1998) and demonstrated by Maitland and colleagues (Maitland, Intrieri, Schaie, & Willis, 2000; Schaie, Maitland, Willis, & Intrieri, 1998) were used to test these relationships. These methods utilize the tau-\(\gamma\) matrix in

### Table 2

**Means and Standard Deviations for All Tests**

<table>
<thead>
<tr>
<th>Variables</th>
<th>35-50</th>
<th>55-65</th>
<th>70-80</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Recognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RN NAMES</td>
<td>7.48</td>
<td>2.37</td>
<td>7.10</td>
</tr>
<tr>
<td>RN FACES</td>
<td>7.26</td>
<td>2.76</td>
<td>6.08</td>
</tr>
<tr>
<td>RN NOUNS</td>
<td>5.27</td>
<td>2.06</td>
<td>4.88</td>
</tr>
<tr>
<td>Recall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC ACTIONS</td>
<td>10.35</td>
<td>2.47</td>
<td>8.43</td>
</tr>
<tr>
<td>RC NOUNS</td>
<td>11.32</td>
<td>2.24</td>
<td>10.01</td>
</tr>
<tr>
<td>RC ACTIVITIES</td>
<td>13.35</td>
<td>3.91</td>
<td>10.31</td>
</tr>
<tr>
<td>RC ITEM INFORMATION</td>
<td>6.61</td>
<td>3.19</td>
<td>5.58</td>
</tr>
<tr>
<td>RC SOURCE INFORMATION</td>
<td>5.73</td>
<td>2.74</td>
<td>4.74</td>
</tr>
<tr>
<td>RC FF*</td>
<td>5.74</td>
<td>1.59</td>
<td>5.31</td>
</tr>
<tr>
<td>RC DF*</td>
<td>4.37</td>
<td>1.21</td>
<td>3.94</td>
</tr>
<tr>
<td>RC DD*</td>
<td>5.15</td>
<td>1.50</td>
<td>4.45</td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOCABULARY</td>
<td>23.19</td>
<td>4.11</td>
<td>22.16</td>
</tr>
<tr>
<td>GENERAL KNOWLEDGE</td>
<td>7.84</td>
<td>1.63</td>
<td>7.85</td>
</tr>
<tr>
<td>Fluency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLUENCY “A”</td>
<td>12.67</td>
<td>4.41</td>
<td>10.72</td>
</tr>
<tr>
<td>FLUENCY “Professions on B”</td>
<td>5.08</td>
<td>2.07</td>
<td>4.87</td>
</tr>
<tr>
<td>FLUENCY “M 5-letter long”</td>
<td>5.89</td>
<td>3.17</td>
<td>5.66</td>
</tr>
</tbody>
</table>

Note. RN = Recognition; RC = Recall. *F = focused attention, D = divided attention; the first letter refers to encoding and the second letter refers to retrieval.
LI.SREL to model the regression intercepts of each observed variable. The alpha and kappa matrices are used to model latent mean deviation values between groups for first- and second-order factors, respectively. An arbitrary reference group is selected; the youngest age group was used for our analyses. In an effort to examine means for the first-order factors, the regression intercepts are allowed to estimate freely whereas the mean values are fixed to zero for the reference group. Intercept values for all remaining comparison groups are held invariant to the reference group (i.e., this tests whether the intercept values from the reference group fit for comparison groups). The group differences for the latent means are expressed in the alpha matrix, as deviation values. Values exceeding the critical level demonstrate significant group differences for that factor mean. Analogous to this process, the second-order factor mean differences are displayed in the kappa matrix in LISREL. The variable intercepts are allowed to freely estimate in the reference group whereas the first- and second-order means are fixed to zero. The first-order means and intercepts are held invariant, whereas the second-order means are freely estimated in all subsequent groups. Significant values demonstrate group differences in the higher order factors.

Whereas these methods provide information about differences between the reference group and each comparison group, they fail to provide a test of differences between comparison groups. Therefore, to evaluate these differences, we used unstandardized parameter estimates and their associated standard errors to compute z tests of pairwise differences (see Hertzog & Bäckley, 2001). This procedure may also be used to examine differences between factor loadings (particularly useful if tests of partial invariance are necessary) and between factor variances and covariances, provided unstandardized estimates are used (e.g., Maitland et al., 2000, 2001).

**Comparative fit statistics.** Model fit was evaluated by examining the following fit indexes: (a) model chi-square; (b) goodness-of-fit index (GFI; Jöreskog & Sörbom, 1989); (c) nonnormed fit index (NNFI; Bentler & Bonnet, 1980); (d) comparative fit index (CFI; Bentler, 1990); (e) root-mean-square error of approximation (RMSEA; Steiger, 1990; Steiger & Lind, 1980); and (f) Z ratio (Boelen, 1989). Explanation and interpretation of these (and additional) fit indexes are available elsewhere (e.g., Bollen, 1989; Browne & Cudeck, 1993; Hu & Bentler, 1995; Kline, 1998).

**Results**

**Unitary Versus Two-Factor Memory Model**

The middle-aged group (35–50 years old) was used for the first series of analyses. We accepted correlated error terms that were theoretically interpretable as shared method variance between observed measures. These correlated error terms were used across all measurement models and included (a) face-name (RN NAMES) with face (RN FACES) recognition, (b) free- (RC NOUNS) recall of self-performed tasks, (c) recall of episodic item (RC ITEM) with episodic source (RC SOURCE) information, (d) all error terms for recall of words under focused (RC FF) and divided (RC DF, RC FD, RC DD) attention, and (e) two forms of letter fluency (FLUA with FLUM).

The single-factor model (M1; Figure 1A) was specified with all 17 observed measures treated as markers of one global memory construct. The configural unitary memory model allowed factor loadings to be freely estimated with the exception of one item set to a value of 1.0 to scale the latent factor. Fit of this model was modest, $\chi^2(109) = 193.28, p < .001$, GFI = .943, CFI = .932. Table 3 provides the fit values for this and all subsequent models. Whereas all factor loadings were statistically significant ($p < .01$), some were weak (e.g., RN NAMES = .35, FLUB = .26). All correlated error terms were significant ($p < .01$) with the exception of RC FD with RC DF. Factor loadings for the completely standardized solution ranged in magnitude from a low of .26 (FLUB) to a high of .66 (VOCAB). Taken together, these results suggested that the fit of the model could be improved.

Next, we tested a model that divided declarative memory into two factors (M2; Figure 1B), with tasks associated with episodic and semantic memory loading onto separate factors. The episodic memory factor was comprised of 12 tests (RN NAMES, RN FACES, RC NOUNS, RC ACTIONS, RC NOUNS, RC ACTIV, RC ITEM, RC SOURCE, RC FF, RC DF, RC FD, RC DD), and 5 tests contributed to the semantic memory factor (VOCAB, GEN KNOWL, FLUA, FLUB, FLUM). The unitary model (M1) is nested within this two-factor model (see Table 3), and M2 showed significant improvement over M1, $\chi^2(108) = 168.69, p = .001$, GFI = .950, CFI = .952; $\chi^2(1) = 24.59, p < .001$. All factor loadings and factor variances and covariances were statistically significant ($p < .01$). All correlated errors for shared method variance were significant ($p < .01$), with the exception of RC FD and RC DF. Standardized factor loadings ranged between .37 and .63 for episodic memory and between .30 and .74 for semantic memory. The correlation between the episodic and semantic memory factors was $r = .79$. These results suggested that the two-factor model fit our data better than the unidimensional model. We next explored whether dividing episodic and semantic memory into 4 first-order factors would further improve model fit.
Comparisons of Declarative Memory Models

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>χ²</th>
<th>p</th>
<th>GFI</th>
<th>RMSEA</th>
<th>NNFI</th>
<th>CFI</th>
<th>Z ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 Unidimensional</td>
<td>109</td>
<td>193.28</td>
<td>.001</td>
<td>.943</td>
<td>.045</td>
<td>.915</td>
<td>.932</td>
<td>1.77</td>
</tr>
<tr>
<td>M2 Two-factor</td>
<td>108</td>
<td>168.69</td>
<td>.039</td>
<td>.950</td>
<td>.039</td>
<td>.940</td>
<td>.952</td>
<td>1.56</td>
</tr>
<tr>
<td>M3 Four-factor</td>
<td>103</td>
<td>149.05</td>
<td>.002</td>
<td>.955</td>
<td>.035</td>
<td>.952</td>
<td>.964</td>
<td>1.45</td>
</tr>
<tr>
<td>M4 Second-order six-factor</td>
<td>104</td>
<td>151.29</td>
<td>.002</td>
<td>.955</td>
<td>.034</td>
<td>.952</td>
<td>.963</td>
<td>1.45</td>
</tr>
<tr>
<td>Age invariance/education covaried</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M5 Configural multigroup</td>
<td>312</td>
<td>359.79</td>
<td>.032</td>
<td>.957</td>
<td>.022</td>
<td>.980</td>
<td>.986</td>
<td>1.15</td>
</tr>
<tr>
<td>M6 Age invariant first-order factor loadings</td>
<td>372</td>
<td>447.91</td>
<td>.004</td>
<td>.940</td>
<td>.026</td>
<td>.972</td>
<td>.977</td>
<td>1.20</td>
</tr>
<tr>
<td>M7 Age invariant first- and second-order factor loadings</td>
<td>376</td>
<td>452.07</td>
<td>.004</td>
<td>.941</td>
<td>.026</td>
<td>.972</td>
<td>.977</td>
<td>1.20</td>
</tr>
</tbody>
</table>

With latent means

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>χ²</th>
<th>p</th>
<th>GFI</th>
<th>RMSEA</th>
<th>NNFI</th>
<th>CFI</th>
<th>Z ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>M8 M7 with second-order means</td>
<td>408</td>
<td>799.15</td>
<td>.001</td>
<td>.941</td>
<td>.056</td>
<td>.859</td>
<td>.875</td>
<td>1.96</td>
</tr>
<tr>
<td>M9 M7 with first-order means</td>
<td>402</td>
<td>557.10</td>
<td>.001</td>
<td>.941</td>
<td>.035</td>
<td>.950</td>
<td>.956</td>
<td>1.39</td>
</tr>
</tbody>
</table>

Note. Z ratio = χ²/df; GFI = LISREL goodness-of-fit index; RMSEA = root-mean-square error of approximation; NNFI = comparative fit index; CFI = normed fit index; M1–M9 = Models 1–9; M-A = middle-aged; Y-O = young-old; O-O = old-old.

* Model M4 was selected as the best fitting of the competing measurement models. 

Four-Factor Model of Declarative Memory

The next hierarchical model (M3; Figure 1C) divided the episodic memory factor into recognition and recall factors. Recognition was comprised of 3 observed tests (RN NAMES, RN FACES, RN NOUNS), and the remaining 9 episodic memory tests loaded onto the recall factor (RC ACTIONS, RC NOUNS, RC ACTIV, RC ITEM, RC SOURCE, RC FF, RC DF, RC FD, RC DD). The semantic memory factor was also divided into two factors: knowledge (VOCAB, GEN KNOWL) and fluency (FLUA, FLUB, FLUM). Model M2 is nested within M3, and fit indexes (see Table 3) revealed a significant improvement over the two-factor model, χ²(103) = 149.05, p = .002, GFI = .955, CFI = .964; χ²/df(5) = 19.64, p < .01. Dividing the previous two factors into four factors improved the factor loading for VOCAB, whereas the loading for RN NAMES on the recognition factor remained similar to previous models. All three fluency measures showed stronger loadings on their own factor. All factor variances and covariances were statistically significant. The pattern of correlated errors was similar to the previous model with one exception: FLUA with FLUM was no longer statistically significant. All factor loadings for the completely standardized solution were significant and ranged in magnitude from a low of .34 (RN NAMES) to a high of .88 (FLUA). One final model remained to be tested within our hypothesized hierarchical structure. This model included episodic and semantic factors, and these were further divided into recall-recognition and knowledge–fluency.

Second-Order Six-Factor Model of Declarative Memory

In the final model (M4; Figure 1D), episodic memory was specified as a second-order factor that influenced performance on the first-order factors of recognition and recall previously described. The second-order semantic memory factor was hypothesized to influence performance on the first-order factors of knowledge and fluency. This model is nested within M3, and the fit for this second-order factor model showed improvement over previously tested models but no significant decrement in fit from M3 (see Table 3): χ²(104) = 151.29, p < .002, GFI = .955, CFI = .963; χ²/df(5) = 2.24, ns. All first-order and second-order factor loadings were statistically significant (p < .01). Correlated errors for shared method variance were significant (p < .01), with the exception of RC FD with RC DF and FLUA with FLUM. Standardized factor loadings ranged from .35 to .43 for the recognition tasks, from .38 to .62 for the recall tasks, from .38 to .85 for the knowledge tasks, and between .34 and .90 for the fluency tasks (see Table 4). The correlation between the episodic and semantic memory factors was r = .80. The nonsignificant decrement in fit between M3 and M4 allowed us to accept M4 as the best-fitting measurement model. The second-order model accounts for the six factor correlations in model M3, using four second-order factor loadings, and the correlation between the two higher order factors. Therefore, M4 provides the most parsimonious representation of the structure of declarative memory in this study. Moreover, given the nonsignificant decrement in fit between M3 and M4, theoretical considerations favor M4 (cf. Byrne, 1998). Specifically, the-
This strategy effectively controlled for differences in educational attainment among age groups. Results showed that the higher order factor model fit well across all groups [M5: $\chi^2(312) = 359.79, p = .032$, GFI$_{middle-aged} = .958$, GFI$_{young-old} = .967$, GFI$_{old-old} = .951$, CFI = .986] and configurial invariance was supported. We then proceeded to test more stringent age invariance models (see Table 3).

A metric model tested age group invariance of all first-order factor loadings [M6: $\chi^2(376) = 452.07, p = .004$, GFI$_{middle-aged} = .948$, GFI$_{young-old} = .957$, GFI$_{old-old} = .940$, CFI = .977; $\chi^2_{diff}(60) = 88.12, p > .05$]. Accordingly, the hypothesis of age invariance of first-order factor loadings was accepted as metric invariance of all factor loadings resulted in nonsignificant decrement in model fit. The next model, a test of structural invariance of second-order factor loadings, constrained all first-order and second-order factor loadings to be equivalent across age groups [M7: $\chi^2(376) = 452.07, p = .004$, GFI$_{middle-aged} = .948$, GFI$_{young-old} = .957$, GFI$_{old-old} = .940$, CFI = .977; $\chi^2_{diff}(4) = 14.16, p > .05$] and resulted in a nonsignificant difference when compared against M6. Therefore, age invariance of first- and second-order factor loadings was accepted. Standardized factor loadings for M7 are provided in Table 4. The correlation between the episodic and semantic factors was $r = .52$ for the middle-aged group, $r = .66$ for the young-old group, and $r = .84$ for the old-old group. Whereas the factor correlations increased across age groups, $z$ tests on the unstandardized estimates revealed no significant differences in their magnitude across the three age groups ($z$ middle-aged vs. young-old = -.68; $z$ middle-aged vs. old-old = -.58; $z$ young-old vs. old-old = -.85).

### Latent Means Model Extensions

The best-fitting covariance models demonstrated measurement and structural equivalence of declarative memory in adults ranging from 35 to 80 years of age. However, these analyses do not yet provide tests of age group differences in level of performance for the six memory factors. Therefore, vectors of means from the observed measures were modeled in addition to the accepted covariance structures to estimate latent mean age group differences (adjusted for educational differences among groups). For the alpha matrices, $t$ values provided by LISREL were used to determine statistically significant age group differences.

Table 5 provides the mean deviation values (alphas), standard errors, and information concerning statistically significant differences in memory performance between the reference group and the two older comparison groups. Additionally, $z$ tests of mean differences between the young-old and old-old groups are provided to aid in interpreting differences between the two older groups. Figure 2 displays latent-mean profiles for the second-order age group differences (M8). The figure includes a confidence band at $\pm 2.58 (p < .01)$ relative to the reference group (middle-aged adults); any group exceeding this band indicates a statistically significant age difference. As can be seen, there was an overall pattern of age deterioration in episodic memory. The young-old age group was significantly better than the old-old group ($z = 4.31$), and both groups were performing significantly worse than the middle-aged group for episodic memory. By contrast, in semantic memory, the young-old group was significantly better than the middle-aged participants, whereas the old-old did not differ significantly from the middle-aged but performed significantly worse than the young-old ($z = 3.79$).

For the first-order comparisons (M9), between-groups variation was tested by setting alpha values for all estimated memory factors to zero for the middle-aged group. Scores for the young-old and

---

**Table 4**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Middle-aged only (M4)</th>
<th>Age invariant (M7)</th>
<th>Education loadings (M7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition</td>
<td>.348</td>
<td>.341</td>
<td>.188</td>
</tr>
<tr>
<td>RN NAMES</td>
<td>.506</td>
<td>.500</td>
<td>.262</td>
</tr>
<tr>
<td>RC ACTIONS</td>
<td>.549</td>
<td>.437</td>
<td>.413</td>
</tr>
<tr>
<td>RC SOURCE</td>
<td>.621</td>
<td>.484</td>
<td>.274</td>
</tr>
<tr>
<td>RC Focused</td>
<td>.384</td>
<td>.419</td>
<td>.216</td>
</tr>
<tr>
<td>RC Semantic</td>
<td>.381</td>
<td>.360</td>
<td>.178</td>
</tr>
<tr>
<td>Knowledge</td>
<td>.419</td>
<td>.326</td>
<td>.242</td>
</tr>
<tr>
<td>VOCAB</td>
<td>.418</td>
<td>.392</td>
<td>.225</td>
</tr>
<tr>
<td>GEN KNOWL</td>
<td>.388</td>
<td>.323</td>
<td>.219</td>
</tr>
<tr>
<td>Fluency</td>
<td>.851</td>
<td>.618</td>
<td>.466</td>
</tr>
<tr>
<td>FLUA</td>
<td>.382</td>
<td>.356</td>
<td>.192</td>
</tr>
<tr>
<td>FLUB</td>
<td>.901</td>
<td>.529</td>
<td>.348</td>
</tr>
<tr>
<td>FLUM</td>
<td>.338</td>
<td>.383</td>
<td>.206</td>
</tr>
<tr>
<td>Production</td>
<td>.738</td>
<td>.513</td>
<td>.275</td>
</tr>
</tbody>
</table>

---

**Note:** All values were statistically significant at $p < .01$. *Educational differences among age groups ($p < .01$). *F = focused attention, D = divided attention; the first letter refers to encoding and the second letter refers to retrieval.
old-old age groups were estimated as deviation values, using the middle-aged group as a reference. It should be noted that intercepts were freely estimated for the middle-aged (reference) group but were constrained to be equivalent among all age groups. Significant age group differences were observed between the old-old group and the middle-aged and young-old age groups for recognition (O-O vs. Y-O; \( z = 3.66 \)), whereas performance of the young-old was not different from that of the middle-aged group (Figure 3). Significant differences among all three groups were found for recall (O-O vs. Y-O; \( z = 4.42 \)). The young-old age group performed better than the middle-aged and old-old groups on knowledge (O-O vs. Y-O; \( z = 2.83 \)), whereas the difference between the old-old and middle-aged groups was not statistically significant. Finally, a marginal difference was noted between the middle-aged and old-old age groups for fluency, although the young-old and middle-aged participants did not differ.

### Discussion

The present results have implications for three fundamental issues: (a) the structure of declarative memory; (b) whether the structure of declarative memory is equivalent across the adult life span; and (c) whether aging affects specific declarative memory functions differentially. These issues are discussed in turn.

#### Structure of Declarative Memory

The model comparisons provided support for a division of declarative memory into episodic and semantic memory and for further subdividing episodic memory into recognition and recall and semantic memory into fluency and knowledge. Although such a fractionation of declarative memory is not entirely novel, to our knowledge few if any previous studies have examined this view in such a comprehensive manner. Therefore, the present findings provide novel support for a division of human long-term memory into multiple systems (see Foster & Jelicic, 1999). Having said that, it should be acknowledged that other interpretations are possible. According to the components-of-processing framework (e.g., Moscovitch, 1992; see also Roediger, Buckner, & McDermott, 1999), some processes are specific to particular memory tasks and others are shared by many tasks. On this view, the loading of a constellation of tasks on a specific factor can be seen as reflecting a shared component process by these tasks. At present, it is not entirely clear what the critical differences are between component processes and memory systems (Parkin, 1999), but for both perspectives it is fundamental to classify measures according to underlying systems or component processes. The present modeling results should be of value for such classification.

#### Age Invariance of the Factor Structure

In both clinical and research settings it is common to contrast the performance of younger and older adults on measures that fall into one or more of the factors that we have identified (i.e., recall, recognition, knowledge, fluency). The validity of group comparisons on a given task depends on the assumption that the underlying structure does not change over the age range of the target groups. If not, the measurements may not represent the same constructs in the age groups. However, few cognitive aging studies have formally tested for age invariance of the relevant factor structure (but see Babcock, Laguna, & Roesch, 1997; Maitland et al., 2000; Schaie et al., 1998), and we know of no test of age invariance for a semantic-episodic factor structure. This is a noteworthy omission because young-old comparisons on measures of semantic and episodic memory populate the empirical literature on cognitive aging and are central to dominant theoretical perspectives within this domain (e.g., Craik & Jennings, 1992; Hultsch et al., 1998; Salthouse, 1991). Our finding of age invariance is therefore of substantial importance for the interpretation of findings of differ-

---

**Table 5**

<table>
<thead>
<tr>
<th>Group</th>
<th>Recognition</th>
<th>Recall</th>
<th>Knowledge</th>
<th>Fluency</th>
<th>Episodic</th>
<th>Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha</td>
<td>(-.155)</td>
<td>(-1.034)</td>
<td>1.099</td>
<td>(.157)</td>
<td>(-.394)</td>
<td>.776</td>
</tr>
<tr>
<td>SE</td>
<td>0.121</td>
<td>0.167</td>
<td>0.315</td>
<td>0.290</td>
<td>0.072</td>
<td>0.258</td>
</tr>
<tr>
<td>(r) value</td>
<td>(-1.281)</td>
<td>(-6.212)</td>
<td>3.488</td>
<td>.542</td>
<td>(-5.460)</td>
<td>3.008</td>
</tr>
</tbody>
</table>

**Old-old age group**

| Alpha          | \(-.880^a\) | \(-2.614^a\) | \(-.302^a\) | \(-.939\) | \(-1.035^a\) | \(-.693^a\) |
| SE             | 0.157       | 0.316  | 0.383     | 0.367   | 0.130    | 0.289    |
| \(r\) value    | \(-5.596\)  | \(-8.272\) | \(-.787\) | \(-2.556\) | \(-7.940\) | \(-2.403\) |

Note. \( r\) values less than ±2.58 are not statistically significant and indicate no age group differences of the latent construct. Positive values indicate better performance, negative values indicate worse performance compared with the reference group (middle-aged).

\(^a\) Indicates statistically significant mean differences between the old-old and young-old groups.
ential effects of aging on episodic versus semantic memory. However, because our youngest age group was middle aged (35–50 years old), we do not know whether the pattern of age invariance generalizes to the age range of young adults (i.e., 20–30 years old) that are so commonly used in cognitive aging research. It should also be noted that the correlation between episodic and semantic memory factors increased with age, although tests of group differences in the magnitude of these correlations failed to reach significance. This trend toward dedifferentiation with increasing age is consistent with previous research (e.g., Babcock et al., 1997; Li & Lindenberger, 1999) and may indicate an emerging change of the underlying factor structure. With these caveats in mind, our observation of age invariance suggests that findings of differential effects of aging reflect specific age differences in level of memory performance within a common factor structure.

Factor-Specific Performance Differences

Comparisons of the second-order latent factor means across groups revealed significant decreases in episodic memory performance with advancing age. By contrast, there was a significant increase in semantic memory from middle age to young-old age, and the old-old group was not impaired relative to the middle-aged group. This observation suggests that aging has a differential effect on episodic and semantic memory. Specifically, age differences are apparent at an earlier age for episodic memory, an observation consistent with many reports of observed mean-level differences (for reviews, see Bäckman et al., 1999; Craik & Jennings, 1992). The demonstration of such a pattern at the level of latent means, in the presence of age invariance at the factor level, increases the strength of the support for such a differential effect.

Figure 2. Age differences for second-order factors. Latent means for the middle-aged group were set to zero and used as a reference group to derive deviation values (expressed as \( t \) values) for young-old and old-old latent means respectively. Negative values indicate lower memory performance relative to the middle-aged reference group. yrs = years.

Figure 3. Age differences for first-order factors. See Figure 2 for details. yrs = years.
At a general level, this differential effect may reflect the fact that increasing adult age is associated with deficits in acquiring and recollecting episodic information, although the ability to use semantic information is relatively little affected by the normal aging process (e.g., Mitchell, 1989; Nyberg, Bäckman, Ernmark, Olofsson, & Nilsson, 1996; for a review, see Bäckman, Small, & Wahlin, 2001). However, analyses at the first-order factor level suggested that age differences within the episodic domain are not homogeneous. Specifically, age differences were more pronounced for recall than for recognition, and the recognition memory performance of the young-old group was not significantly different from that of the middle-aged group.

The finding that recall was more age sensitive than recognition is concordant with theoretical arguments that age-related differences in memory performance should be reduced as the level of cognitive support increases, because of age-related deficits in self-initiated processing and recoding operations (Bäckman, 1989; Craik, 1983). Previous research has yielded only partial support for this view. In particular, although single studies have demonstrated selective gains from the provision of cognitive support in both older (e.g., Bäckman & Larsson, 1992; Craik & McDowd, 1987; Sharps & Gollin, 1988) and younger (Kliegl, Smith, & Baltes, 1989; Puglisi & Park, 1987; Rabinowitz, 1989) adults, the typical finding is one of parallel gains from cognitive support across the adult life span (e.g., Bäckman, 1991; Cohen, Sandler, & Schrock, 1987; Dixon & Gould, 1998; Johnson, Schmitt, & Pitruko wicz, 1989; Rabinowitz & Craik, 1986; for a review, see Bäckman et al., 1990; Light, 1991). Given the lack of consistency in research findings concerning the relationship among aging, cognitive support, and episodic memory performance, it is interesting to note that the present findings are consistent with age-differential effects of cognitive support in latent factors, using multiple measures of both recall and recognition.

In the context of factor-specific performance differences, there was also some variation within the semantic domain. Specifically, the performance on knowledge-based tests improved into young-old age and showed little decline in old-old age relative to middle age. In contrast, fluency performance showed minimal improvement in young-old age and tended to decrease significantly in old-old age.

In the psychometric literature on intelligence, it has been demonstrated that knowledge-based tasks, such as vocabulary, may show increases in middle adulthood and stability up to very old age (e.g., Kaufman, 1990; Schaie, 1996). These findings are corroborated in the experimental literature, showing that semantic knowledge may accumulate across the life span (e.g., Light, 1992; Salthouse, 1988) and that the organization and associative structure of the semantic network remains stable into old age (e.g., Balota & Duchek, 1988; Burke & Peters, 1986; Laver & Burke, 1993). However, the latter literature also demonstrates that aging is associated with increasing difficulties in retrieving lexical information at a rapid rate (for a review, see Light, 1992). Age-related slowing of lexical access has been demonstrated in numerous paradigms assessing word-finding ability (e.g., Burke, MacKay, Worthley, & Wade, 1991; Crook & West, 1990; Maylor, 1990). Thus, the present finding of a dissociative pattern of knowledge and fluency from middle to very old age is consistent with the bulk of related research. As with the results on episodic recall and recognition, it is notable that the current results on semantic knowledge and fluency were observed in multiply-marked latent factors rather than for observed means from single tests as is often reported in the literature.

**Conclusion**

By supporting a multifactor model comprised of second-order factors for episodic and semantic memory and first-order factors for recall–recognition and knowledge–fluency, the present modeling results suggest a differentiation of declarative memory functions at the level of specific factors. The tests of age group differences in memory performance for the latent factors provided additional support by showing that aging does not impair all declarative memory functions equally. Rather, aging had a pronounced negative effect on recall, a smaller effect on recognition, a weak effect on fluency, and a positive effect on knowledge up to age 65. The observation of selective effects of aging on declarative memory functions is highly meaningful in view of the fact that the identified factor structure was age invariant. Hence, the structure of declarative memory was found to be equivalent across the entire age span studied, but within this structure aging had differential effects on level of memory performance.

**References**


Stability, Improvement, and Decline in Adult Life-Span Development of Declarative Memory: Cross-Sectional and Longitudinal Data from a Population-Based Study

Michael Rönnlund¹, Lars Nyberg¹, Lars Bäckman²,³, and Lars-Göran Nilsson⁴

¹Department of Psychology, Umeå University, ²Department of Psychology, Uppsala University,
³Stockholm Gerontology Research Center and Division of Geriatric Epidemiology, NEUROTEC, Karolinska Institute, ⁴Department of Psychology, Stockholm University

Address correspondence to:
Michael Rönnlund
Department of Psychology,
Umeå University,
S-901 87 Umeå,
Sweden
E-mail: michael.ronnlund@psy.umu.se

Abstract
Five-year changes in episodic and semantic memory performance were examined in a sample of 829 participants from 10 age cohorts (35-80 years). A second, cohort-matched, sample (n = 959) was assessed at Time 2 to control for practice effects. Longitudinal changes indicated improved episodic memory performance for the younger cohorts, in contrast with the gradual age-related decrements depicted by cross-sectional data. For groups 60 years or older performance declined substantially. Practice effects were observed for the episodic measure, but even when these were adjusted for results were consistent with a late onset (> 60 years) of decline. Semantic memory showed minor improvements until age 55, with less steep decrements in old age as compared with episodic memory. Cohort differences in educational attainment appear to account for discrepancies between cross-sectional and practice-adjusted longitudinal data, both for episodic and semantic memory. Collectively, the results show that the age-trajectories for episodic and semantic differ, and underscore the need to control for cohort and retest effects in cross-sectional and longitudinal studies, respectively.

Key Words: Episodic Memory, Semantic Memory, Cross-sectional, Longitudinal, Cohort Differences, Practice Effects
The bulk of evidence concerning adult cognitive development emanates from cross-sectional studies. This is so despite the potential pitfalls associated with this design, especially the inherent confounding of effects of maturational change and those of cohort membership. Indeed, substantial improvements have been observed on standard tests of intellectual functioning over the last century, for example at time of recruitment into military service (Flynn, 1987; Neisser, 1998; Rodgers, 1999). Such time-lag gains together with other systematic evidence of positive cohort gradients (Schaie, 1994, 1996) suggest that age differences in cognitive test performance observed at a single point in time may overestimate the magnitude of true age change due to cohort-related influences.

In the present study, adult life-span development of two forms of declarative memory was examined: memory of specific events (episodic memory) and memory of general world knowledge (semantic memory). Confirmatory factor analyses have shown that episodic memory and semantic memory constitute separable ability factors (Nyberg, 1994; Nyberg et al., 2003), in line with various task dissociations (Nyberg & Tulving, 1996). We next review past cross-sectional and longitudinal evidence of aging patterns on measures of these constructs.

Aging Patterns in Episodic Memory Measures

Episodic memory is typically regarded as highly age sensitive, a notion primarily based on cross-sectional extreme age-group comparisons (for reviews, see Bäckman, Small, & Wahlin, 2001; Kausler, 1994; Smith, 1996; Verhaeghen & Salthouse, 1993). Studies covering a wider age range have confirmed and extended this finding. These studies indicate that onset of decline may occur as early as in the 20s or 30s, and that individuals perform as much as one standard deviation (SD-) unit below peak level when they reach their 60s. At age 80, the size of this effect may approach 2 SD-units (e.g., Nilsson et al., 1997; Schaie, 1994, Park et al., 2002; Verhaeghen & Salthouse, 1997). Thus, the deterioration of episodic memory functioning appears to be approximately linear.
Several longitudinal studies appear to contradict the notion that episodic memory declines gradually across the adult life span. A number of studies have found no time-related changes within groups of middle-aged or old participants in one or several episodic measures (e.g., Flicker, Ferris, & Reisberg, 1993; Hasker & Klebe, 2001; Hultsch, Hertzog, Small, McDonald-Miszcza, & Dixon, 1992; McDonald-Miszcza, Hertzog, & Hultsch, 1995). In other cases, time-related improvements have even been observed (e.g., Bäckman et al., 1998; Flicker et al., 1993; McCarty, Siegler, & Logue, 1982; Hultsch et al., 1992, Mitrushina & Satz, 1991). As noted by Salthouse (2000), lack of power of detecting decline may be a factor in longitudinal studies, but time-related improvements in performance are obviously difficult to reconcile with cross-sectional findings.

However, other longitudinal studies have reported time-related decline (e.g., Christensen, 2001; Johansson, Zarit, & Berg, 1992; Mayeux, Small, Tang, Tycko, & Stern, 2001; Shichita, Hatano, Ohashi, Shibata, & Matuzki, 1986), often using longer retest intervals relative to the former set of studies. Moreover, a recurrent pattern in studies covering a wider age range is that baseline age is negatively related to memory change (e.g., Arenberg, 1982; Colsher & Wallace, 1991; Finkel, Pedersen, Plomin, & McClearn, 1998; Zelinski, Gilewski, & Schaie, 1993; Zelinski & Stewart, 1998). That is, with passage of time, young/middle-aged groups remain stable, or even improve, whereas reliable and positively accelerated decline is observed for older groups.

Seven-year data from the Seattle Longitudinal Study (SLS; Schaie, 1994) on “verbal memory” (a factor score primarily reflecting two episodic word recall measures) are illustrative of the foregoing pattern. In the period from young adulthood (25-28 years) to midlife (the 60s), minor improvements were observed and no detectable deficit relative to the mean for the youngest was found until age 67. Past this age, pronounced decline was observed. Similarly, Zelinski and Burnight (1997) observed no 16-year changes in list recall for a group initially 30-36 years old, whereas older groups declined at a rate similar to that predicted from cross-
sectional data. If valid, these results indicate that the notions of an early onset and linear progression of decline of episodic memory functioning from early to late adulthood, indicated by cross-sectional data, may not provide an accurate description of maturational change.

One possible reason for the discrepancy between the cross-sectional and longitudinal patterns is that cohort differences serve to create the impression of a gradual decline in the former case. The secular increase in quantity of educational attainment during the 20th century is a possible factor. This increase tends to be more pronounced for those born later than for those born earlier in the century (e.g., Smith, 1993). Education is related to episodic memory (e.g., Lövdén, Rönnlund, & Nilsson, 2002; Nilsson et al., 1997) and part of the correlation is likely due to a causal influence from education to performance. Thus, cohort-related differences in education possibly account for the tendency that cross-sectional and longitudinal patterns differ for young/middle-aged participants rather than for older participants.

Aging Patterns in Semantic Memory Measures

Semantic memory, reflected by performance on tests of vocabulary, fact recall, and comprehension, also known as tests of crystallized intelligence (Horn & Cattell, 1966, 1967), often exhibits a radically different age-related pattern than episodic measures (for reviews, see Bäckman et al., 2001; Kausler, 1994). In the young to young-old period, cross-sectional data indicate stability of, or increase in, performance (e.g., Birren & Morrison, 1961; Wechsler, 1981). In the period from young-old age to late senescence some studies report negligible differences (e.g., Christensen, 2001; Park et al., 2002) whereas others (e.g., Berkowitz, 1953; Lindenberger & Baltes, 1994) are suggestive of a deficit in very old age. Cross-sectional data from representative samples may be suggestive of an earlier onset of decline. However, adjustment for education reveals a convex age-related trend in this case, consistent with small improvements from age 35 to age 65, followed by late deficit (Bäckman & Nilsson, 1996; Kaufmann, 2001). The latter aging trends are highly similar to those discernible from longitudinal studies of verbal meaning (e.g., Schaie, 1994; Schaie & Strother, 1968a). Thus, for
semantic memory, cross-sectional and longitudinal patterns may be expected to converge once education is adjusted for (Horn & Hofer, 1992).

Thus, from the point of the literature it appears fairly well established that episodic and semantic memory measures are associated with differential aging patterns, with earlier onset and more pronounced decline in old age for episodic memory. However, few studies have included measures of both constructs (e.g., Mitchell, 1989; Nyberg et al., 2003). This holds in particular for longitudinal studies. Interestingly, data from the SLS (Schaie, 1994, 1996) demonstrated highly similar longitudinal age trends for episodic and a semantic measures (verbal memory versus verbal meaning). This and other deviations from the aforementioned age differential pattern (e.g., Flicker et al., 1993), seem to stem from a relative lack of time-related episodic decline (or improvement), as discussed previously.

**Attrition and Practice Effects**

As should be evident from the above review, discrepancies between cross-sectional and longitudinal data on changes in declarative memory measures exist in the literature. Cohort factors that distort the cross-sectional estimates may be hypothesized to account for such discrepancies. However, longitudinal data are themselves associated with internal validity threats that challenge such a conclusion (e.g., Hertzog, 1996; Salthouse, 2000; Schaie, 1988). Two such threats are attrition (i.e., dropout or experimental mortality) and practice effects.

A common finding in longitudinal studies of memory and other cognitive functions is that those who dropped out during the retest interval performed worse at baseline than those who returned (e.g., McCarty et al, 1982; Zelinski & Burnight, 1997). This fact probably reflects a variety of factors, including social and health-related conditions (e.g., Powell et al., 1990). Thus, there may be multiple reasons for not showing up at retest (unwillingness to participate, death, moved from the area, etc.). The magnitude of the attrition effect is often small to moderate (0.2-0.5 SD) and appears to be age invariant (e.g., Baltes, Schaie, & Nardi, 1971; McDonald-Miszczak et al., 1995, Schaie, 1996).
As a result of such selective attrition, the returnee data likely overestimate performance levels within the target population. The extent of dropout is crucial in this regard; a given retest status difference should have a very different mean-level impact depending on whether 10% or 50% of the participants drop out. Even if the attrition effect is similar across age, older participants are more likely to drop out due to mortality or illness. As a result, follow-up data may reveal attenuated or even reversed age differences (e.g., McCarty et al., 1982; Perls, Morris, Ooi, & Lipsitz, 1993; Bäckman, Small, Wahlin, & Larsson, 2000). However, findings that attrition tends to be negatively selective does not necessarily imply that estimates of time-related change are biased. In fact, attrition Initial ability level bear little, if any, relation to rate of memory change from viewpoint of several studies (e.g., Kausler, 1991; Schaie, 1996). Thus, when attrition is kept at fairly low levels and the participants are healthy and screened for dementia, attrition may be expected to a minor effect on mean-level changes.

Words of caution that prior exposure to the test-battery may have affected retest-performance are often provided in discussions of longitudinal results. However, the magnitude of practice effects and their boundary conditions are largely unknown. In general, practice effects may be expected to be larger with a shorter test-retest interval. In line with this assertion, studies reporting time-related improvements often used a relatively short retest interval (< 2 years; e.g., Hultsch et al., 1992; Bäckman et al., 1998). Also, estimates of change (across a given age span) may be more negative with a longer retest interval (see e.g., Schaie, 1996), something which possibly reflect vanishing influence of practice with passage of time.

If practice effects operate in longitudinal studies, it is essential to know if they are age-related. To the extent that recollection of details from the prior test occasion (materials, instructions, etc.) may be a factor, a finding of more pronounced effects for younger adults is not unreasonable. If this was the case, longitudinal indications of a late episodic memory decline (Schaie, 1994; Zelinski & Burnight, 1997) could simply be a result of selective practice effects (i.e., practice effects may serve to mask a true linear age-performance relationship).
One way of estimating the magnitude of practice effects is to compare the mean-level performance in the group of returnees with that of a cohort-matched and previously untested group of participants (Schaie, 1965, 1988). For this comparison to be valid, the original and new sample should (a) be equally representative of the target population and (b) large enough to keep random sampling error at a minimum. In addition, the mean-level impact of attrition must be taken into account with regard to retest performance in the longitudinal sample. This effect may be estimated from Time 1 data. On the basis of the mean difference between previously tested and untested groups and the estimate of the attrition effect, practice-related effects may be sorted out (Schaie, 1988). A step-by-step-description of this method is provided in the results section.

Prior analyses of the influence of practice effects on cognitive test performance (Schaie, 1988) revealed fairly small effect sizes (0.1 to 0.2 SD). In that study, the retest interval was seven years, however. Moreover, the overall patterns of change (cumulative change scores) were quite different depending on whether the practice effects were adjusted for or not. Specifically, whereas the unadjusted means indicated improvements from age 27 to age 67 in verbal meaning, a plateau in performance was apparent when practice effects were adjusted for (Schaie, 1988, Figure 8.1; see, also Horn & Donaldson, 1976; Salthouse, 1991). Thus, practice effects may distort longitudinal data (Salthouse, 1991, 2000), and it is important to control for them, especially in light of the surprising lack of time-related decline, or even improvements, observed in several longitudinal studies of memory.

In the context of the present study it is warranted to comment on some additional methodological issues. Following Schaie (1965), a standard method used to contrast cross-sectional age differences and longitudinal changes has been to compare the F-ratios of Cohort (or age) and Time effects in a so called cross-sequential (i.e., Cohort x Time) (M)ANOVA. However, as argued by Botwinick and Arenberg (1976; see also Adam, 1977), disparate time spans for the cohort and time factors (45 versus 5 years in the present case) may bias the result in
favor of a larger main effect (F-ratio) for the Cohort factor. Thus, such analyses should be supplemented with analyses of the predicted slopes of age change derived from the cross-sectional and longitudinal data. One means to illustrate and allow for comparisons of the predicted losses and gains across the studied age range is to plot the mean cumulative change scores against the cross-sectional data (Schaie, 1994, 1996; Schaie & Strother, 1968a, b).

If data are available for two independent cohort-matched samples tested at various times of measurement another type of cross-sequential (Cohort X Time) analyses may be conducted. In this case, the Time 1 data for the entire original sample (Sample 1) is contrasted with the Time 2 data for the new sample (Sample 2), treating time as a between-, rather than a within-subjects factor (as in the cross-sequential analyses mentioned previously). Such analyses are sometimes discussed under the heading of longitudinal methods (Menard, 1991; Schaie and Strother, 1968b), but are perhaps best considered quasi-longitudinal (Horn & Donaldson, 1976) as they involve a repeated cross-sectional comparison. As we shall see later, these analyses are closely related to the practice–adjustment procedure outlined above.

The main criticism directed at early studies (Schaie & Strother, 1968b, Schaie, Labouvie, & Buech, 1973) that employed independent sample analyses (e.g., Horn & Donaldson, 1976) regarded the interpretation by the authors that the time-related effects were similar to those observed in the corresponding within-subjects analyses. In fact, mean trends were suggestive of magnified time-related decline as determined from the between-sample contrast (see also Salthouse, 1991, 2000). Another objection was that the within-sample data suffered from an overall dropout rate of about 40%. Extensive attrition makes differences in outcomes between the within- and between-sample analyses difficult to interpret; they may result from attrition rather than differential practice (i.e., the effects are confounded). Of additional concern, Horn and Donaldson (1976) argued that differences in F-ratios for the Time factor between the cross-sequential analyses within (repeated measures ANOVA) and between samples (mixed
Adult memory development

ANOVA) should be interpreted with caution; the fact that the time factor varies between subjects in the latter case possibly bias the results towards a smaller time-related effect.

The Present Study

The objective of the present study was to trace the course of episodic and semantic memory development using both cross-sectional and longitudinal data. To this end, data from the first two measurement occasions of a large-scale sequential population-based study of memory and health (the Betula Study, Nilsson et al., 1997) were analyzed. The longitudinal sequence involved 829 participants from ten age cohorts (35 - 80 years old initially) tested five years apart. To control for potential practice effects, data for ten previously untested cohort-matched groups were examined in conjunction with the repeated-measures data. The influence of cohort-related differences in level of schooling on potential discrepancies between cross-sectional and longitudinal results was paid especial attention to. The analyses were conducted at the level of factor scores, using a similar set of indicators as in a model recently found to be metrically age invariant (Nyberg et al., 2003).

Method

Participants

The sample included at the first time of measurement, Sample 1 (S1) at Time 1 (T1), involved 1000 participants. These were from ten age cohorts: 35, 40, 45, 50, 55, 60, 65, 70, 75, or 80 years at date of test. The data collection started in 1988 and was completed in 1990. Date of birth within each of the ten cohorts consequently varied within a two-year range (Mean birth year: 1909, 1914, …, 1954). Each age group comprised 100 participants. The participants were recruited by random sampling via the population registry in Umeå, a community in Northern Sweden with about 100 000 inhabitants. Individuals were excluded if they: (a) suffered from severe visual or auditory handicaps, (b) had a dementia diagnosis (c) suffered from mental retardation, or (d) had another language than Swedish as their native tongue. For further information concerning the recruitment procedure, the reader may consult Nilsson et al (1997).
Comparisons between participants and non-participants (initially contacted persons unwilling to participate) and between participants and the Swedish population on various background variables indicated no differences between participants and refusals (Nilsson et al., 1997), suggesting that the final sample was representative of the target population. Slightly higher levels of education as compared with Sweden as a whole were noted for the youngest cohorts, something which likely reflects the fact that Umeå was established as a major university-city in the 1960s.

Five years after the initial testing (1993-1995), those participants who were still alive (n = 940; 60 were deceased) were contacted again. Of the contacted persons, 48 (5.1%) refused to participate further. Four individuals were unable to participate due to the fact that they had moved from the area. In addition, 16 persons could not be contacted or were unable to turn up for retest due to illness. Thus, a total of 875 of the S1 participants took part in the follow-up assessment. Seventeen of the returnees were diagnosed as demented, and were therefore excluded from the present analyses. For 29 of the remaining participants, T2 data for the cognitive variables of interest were completely missing. This category included participants who had moved, but were willing to complete some survey material concerning health status and social variables via telephone. Finally, for a few individuals (n = 6), the T2 data for the memory tasks were incomplete and dropped from the analyses. Thus, the longitudinal analyses included data for 829 returnees (82.9% of the original sample). For the present purposes, the remaining participants (17.1%) will be referred to as dropouts, regardless of reason for failing to complete the retest. A summary of select participant characteristics across age group and retest status is given in Table 1. In addition to gender distribution and mean self-reported number of years of formal education, values on two cognitive background variables are included: Mini-Mental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975) and Block Design (Wechsler, 1981).

A second sample (S2) was recruited and tested for the first time at T3. The recruitment procedure and the inclusion criteria were the same as for S1. These participants were
drawn from the same birth cohorts as were those in S1 (i.e., mean birth year; 1909, 1914, …, 1954) and were, thus, 40-85 years old at Time 2. S2 included 966 participants; the initial plan to include 100 individuals per cohort (Nilsson et al., 1997) was not feasible for the oldest groups as there were not enough individuals within the target area who met the inclusion criteria and were willing to participate. For 7 of the S2 participants, the memory data were incomplete. These cases were discarded from the following analyses. Thus, the final number of participants included from S2 was 959. Participant characteristics for S2 are provided in Table 2. In addition to the variables that were presented for S1, a measure of perceived memory change (participants rated how they judged their memory to be today as compared with five years ago on a five-point scale; 1 = much worse; 5 = much better) is included. The corresponding mean values for S1 at T2 are provided for the cognitive measures for sake of the sample comparison presented below.

**Dropout Analysis**

Inspection of the values in Table 1 shows that dropout rates were low (about 10 %) across the seven youngest age cohorts. For the groups 65 years and older at baseline, dropout was somewhat higher, for natural reasons (i.e., discarding cases where data were missing due to death or dementia the rates were similar across age). Even for the 80-85 year group, the dropout did not exceed 40%. Out of the total number of dropouts, 88 were women and 83 were men, frequencies that did not deviate from those expected from the original sex distribution, $\chi^2(1) = .336, \text{n.s.}$

For the remaining variables, Retest status X Age cohort analyses of variance (ANOVAs) were conducted. Due to the small $n$’s for dropouts in the younger/middle-aged groups in particular, adjacent groups were collapsed so as to form five groups (35-40, 45-50, 55-60, 65-70, 75-80 years old at T1).

The ANOVA conducted on years of formal education revealed no effect of Retest status, $F(1,990) = 1.67, p > .10$. However, the analyses of the cognitive background variables revealed somewhat higher means for returnees; for MMSE, $F(1,990) = 14.41, \eta^2 = .014$ (mean
for dropouts = 27.35; mean for returnees = 27.95), and for Block Design, $F(1, 980) = 24.07, \eta^2 = .024$ (mean for dropouts = 22.58; mean for returnees = 26.53). No interactions between Retest status and Age were obtained (all $p$s >.20). A highly significant effect of Age cohort was observed for each of the foregoing variables; education, $F(4,990) = 54.96, \eta^2 = .182$; MMSE, $F(4, 990) = 26.68, \eta^2=.104$; and Block Design, $F(4, 990) = 66.61, \eta^2=.212$. In the case of educational attainment, it should be noted that the cohort-related increase is clearly non-linear. More specifically, the means in Table 1 show that the increase from groups born in 1909 to 1934 was about one year. By contrast, there was almost a six-year increase in mean formal schooling across the five consecutive birth cohorts (i.e., those born 1934-1954).

In summary, dropout rates were low in general, but higher for the oldest age cohorts. With respect to sex and education, attrition was non-selective. Dropouts were found to perform somewhat worse, on average, than returnees on the cognitive background variables (MMSE and Block Design). However, at follow-up assessment, the positive bias with regard to mean level performance was likely very small due to the low dropout rates.

Sample Comparison on Background Variables

The validity of the method of comparing independent samples rests on the assumption that the sample composition is equal. The present use of random sampling should likely yield comparable groups. Nevertheless, the assumption should be tested. For this purpose, comparisons of participants in S1 and S2 were conducted for the variables included in the attrition analyses (i.e., sex, education, MMSE and Block Design scores), and for a measure of perceived memory change.

The data in Table 2 show that the sex distribution in S2 was similar to that in S1, with slightly more female participants overall. Also, mean years of formal schooling were comparable in both samples across the cohorts. Although attrition rates in S1 were low, there was evidence of negatively selective attrition for MMSE and Block Design. Thus, one might expect a small advantage for the returnees in S1 over S2 participants for these measures at the
second measurement occasion. A trend in this direction is indeed discernible from the means in Table 2, although the difference is very small. Finally, the measure of perceived memory change appears to vary minimally across samples as judged from the means.

The Sample X Age cohort ANOVA on years of schooling revealed no main effect of Sample. However, the analyses of the MMSE data yielded a significant effect of Sample, consistent with a minimal, but reliable, advantage for returnees in S1 over participants in S2, $F(1, 1768) = 7.72, \eta^2 = .004$. Also, the ANOVA on the Block Design data revealed a significant, but again, very small effect of Sample, $F(1, 1766) = 9.43, \eta^2 = .005$. These effects are small enough to be accounted for by the presence of selective attrition from S1 (see Table 1). No interactions between the two factors were observed ($p_s > .20$). Finally, the ANOVA on the measure of perceived memory change showed no effect of Sample, but of Cohort, $F(9, 1760) = 14.40, \eta^2 = .069$, consistent with increased perceived decline with advancing age. The effects of Cohort were very similar to the ones reported previously as regards the other measures.

Concerning education, the pattern of a magnified cohort-related increase from the early-born to later-born cohorts (1909-1954) was virtually identical to that observed for S1.

To summarize, participants in the new sample (S2) did not differ from the original sample (S1) with regard to years of formal education, the samples was comparable with respect to sex distribution and the participants in the two samples reported similar levels of perceived memory change at Time 2. The $T_2$ advantage of S1 over S2 participants observed for MMSE- and Block Design scores were minimal as judged from effect-size measures, and may result from small, but selective, attrition in S1 and/or practice effects. On the basis of these analyses it appears justified to regard the samples as very similar across the included variables.

Memory measures
Descriptions of the memory measures, five episodic and five semantic memory measures, included in the forthcoming analyses are provided below. As noted, the cognitive data were collected during two test sessions, both of which lasted for 11/2-2 hours. The specific test procedures have been presented in more detail elsewhere (Nyberg et al., 2001; see Nilsson et al., 1997 for a full description of the Betula battery). Participants were requested to use corrective lenses or hearing aids, if used, during testing.

**Episodic Measures**

**Action recall.** Immediately following presentation of 16 verbal commands (e.g., point at the book, lift the cup) that were enacted by the participants (e.g., Cohen, 1981) participants were requested to recall as many of them orally, in any order. In the control condition, 16 sentences were presented without enactment. Number of sentences recalled (correct verb and noun) in the enacted condition was entered in the present analysis.

**Recall of verbal and action nouns.** Following a brief retention interval participants were asked to recall as many nouns as possible from the enacted/nonenacted sentences earlier. The four categories to which each noun belonged served as cues to remember the nouns. Number of nouns recalled from the enacted and nonenacted sentences served as separate measures in the present analyses.

**Recognition of verbal nouns.** Approximately 60 minutes into the second test session, participants were presented with a list of 32 nouns. Half of these were from the enacted/nonenacted sentences studied previously, with eight nouns from each condition (the 16 other nouns were distractors). Number of hits minus false alarms for nouns in the nonenacted condition was adopted as the dependent measure in the present analyses.

**Statement recall.** Participants were asked 20 questions (e.g., ‘What does Astrid Lindgren collect as a hobby?’) that corresponded to 20 fictitious statements about famous and non-famous people studied earlier in the test session. These questions were intermixed with 10 questions that were only possible to answer on basis of prior knowledge and 10 questions about
famous and non-famous persons that could neither be answered on basis of prior knowledge nor
statements presented within the test session (i.e., these served as foils).

Semantic Measures

General knowledge. The items in this test appeared as part of the last of the
episodic measure above (Recall of recently acquired facts). Ten questions included in this test
were possible to answer on the basis of acculturated knowledge (e.g., “Where was Ingemar
Stenmark born?”). Number of correctly answered statements within this category served as the
outcome measure.

Vocabulary. A 30-item multiple-choice synonym test (SRB; Dureman, 1960) was
used. The task was to select a synonym of the target word from among five alternatives. Seven
minutes were allotted for completing the test. The total number of correctly identified synonyms
was entered into the analyses.

Word fluency. Three tests, in which the participants generated as many words as
possible in one minute, were used. In the first test, the participants said aloud as many words as
possible with an initial letter A. The second test involved producing as many words as possible
beginning with the letter M and containing five letters. The third test required generating as
many professions as possible with the initial letter B.

Estimation of Two-factor Model

The memory measures above were used to estimate a two-factor model (i.e.
episodic, semantic) using AMOS 4.0 (Arbuckle & Wothke, 1999). We allowed for correlated
error terms, including free with cued recall of actions, cued recall with recognition of nouns from
sentences, and two measures of letter fluency (Fluency A with Fluency M). Data for the
complete S1 at T1 (n = 1000) were used. The results indicated an acceptable fit of the model as
judged from standard criteria, $\chi^2 = 81.06$, df = 31, $p < .001$, RMSEA = 0.04, GFI = .984. The
correlated error terms were significant (.36, .35, and .30, respectively, for the action, and FLU
measures, $p < .01$). The factor loadings for episodic and semantic memory were all significant and
ranged between .64 and .74, and between .83 and .51 for the episodic and semantic measures, respectively. Factor scores for T₁ and T₂ were computed by summing the products of the (standardized) variables and the factor-loadings. A two factor-model including the present set of indicators is time-invariant in the sense that factors loadings do not differ significantly across times of measurement (Lövdén et al., 2003).

Results

The episodic and semantic factor scores were scaled in t-score units on basis of Time 1 data for the complete Sample 1. Means for the factor scores and the individual memory measures are presented in Tables 3 (episodic memory) and 4 (semantic memory) by Age cohort and Time of measurement. At the level of factor scores, Time 1 data are presented for the complete sample by retest status (dropouts, returnees). All of the subsequent analyses involved the factor scores. Stability coefficients were .78 for the episodic factor, and .85 for the semantic factor. These values indicate considerable stability of individual differences in memory performance across the retest interval (cf., McDonald-Miszczak et al., 1995).

First, the results of the attrition analyses for the episodic and semantic memory factors will be presented. Next, the results of cross-sequential analyses of Time 1 and Time 2 for the returnees are reported for the memory measures. Thereafter, an examination of the influence of practice effects on the longitudinal (within-subjects) data is reported. Finally, the influence of differences in educational attainment on differences between the cross-sectional and longitudinal estimates of change is examined. The alpha-level was set to 0.01, given the large sample sizes.

Attrition Analyses for the Memory Measures

In two Retest status X Age ANOVAs, one for the episodic measure and one for the semantic measure, the T₁-data of those who were retested at T₂ were compared with corresponding data of those who did not return. As with the previous dropout analyses, five levels were used for the age cohort factor. Inspection of the means in Tables 3 and 4 indicate that mean-level performance differed as a function of retest status for both memory factors.
Specifically, a numerical advantage in favor of the returnees is present in 19 of the 20 contrasts (i.e. across all age groups in the case of the semantic measure, and for all but one group in the case of the episodic measure). The mean difference amounts to about 4 t-score units for each of the two measures. As judged from the means in Table 3, there is no apparent age trend with regard to the effect size. Consistent with these observations, the ANOVAs yielded significant effects of Retest status; for episodic memory, $F(1, 990) = 24.62, \eta^2 = .024$ (for returnees, mean $t$ score = 50.79; for dropouts, mean $t$-score = 46.98), and for semantic memory, $F(1, 990) = 21.25, \eta^2 = .021$ (for returnees, mean $t$ score = 50.62; for dropouts, mean $t$-score = 46.75). No interactions between Retest status and Age were observed (all $p$s > .20), indicating that the pattern of selective attrition was age invariant.

**Cross-Sequential Analyses**

The returnee data for the episodic and semantic memory measures were next subjected to separate Age/Cohort X Time ANOVAs in order to contrast cross-sectional (Age) and longitudinal (Time) effects (Schaie, 1965, 1977).

**Episodic Memory**

Returning to Table 3, and focusing on the data for returnees, advanced age was clearly associated with lower levels of episodic memory performance at both times of measurement. At odds with these patterns of data, a clear tendency for time-related improvements was observed for groups ranging in age from 35 to 55 years at baseline. By contrast, the older groups mean-level performance was worse at the time of retest, a tendency that is progressively more noticeable with advancing age. The results of a 10 (Cohort) X 2 (Time) ANOVA revealed significant effects of Cohort, $F(9, 819) = 49.83, \text{MSe} = 121.60, \eta^2 = .354$, and of Time, $F(1, 819) = 7.78, \text{MSe} = 21.86, \eta^2 = .01$, reflecting worse performance with advancing age, and lower performance at follow-up measurement, respectively (the size of the latter effect was very small). Consistent with the observation that the direction of time-related
influences were positive for the youngest groups but negative for the older groups, the Cohort X Time interaction was highly significant, $F(9, 819) = 11.03$, $MSe = 21.86$, $\eta^2 = .108$.

**Semantic Memory**

The means for returnees for the semantic measure, presented in Table 4, also reveal a negative cross-sectional age trend at both times of measurement. As for the episodic measure, the time-related changes were positive for the groups in the age range 35 to 50 years. Also, there is a clear trend of a time-related deficit for the older participants, the 80-year-old group in particular, although the magnitude of decline, as discernible from the mean change scores, is not as pronounced as for the episodic measure (See Table 3).

The $10$ (Cohort) X $2$ (Time) ANOVA on the semantic memory data yielded significant main effects of Cohort, $F(9, 819) = 24.39$, $MSe = 150.93$, $\eta^2 = .211$. No effect of Time was observed, $F(1, 819) = 1.34$, $p > .20$, but, as in the analyses of episodic memory data, the Cohort X Time interaction was significant, $F(9, 819) = 7.07$, $MSe = 13.38$, $\eta^2 = .072$.

To summarize, the cross-sequential analyses indicated a very different picture of developmental patterns in episodic and semantic memory, depending on whether cross-sectional (Cohort) effects or longitudinal (Time) effects were examined. Specifically, the cross-sectional age differences were quite substantial for both episodic and semantic memory, as visible from the mean patterns and as indexed by strong cohort effects. By contrast, time-related improvements were observed across the younger age cohorts for both declarative memory factors. This was most clearly evident for the episodic memory measure, a finding that is difficult to reconcile with the common notion of an early onset of decline for episodic memory. In fact, the mean change scores in Table 4 (episodic factor) and Table 5 (semantic factor) reveal quite similar time-related patterns.

In Figure 1, composite cross-sectional age gradients for the episodic and semantic memory factors are presented. These were based on the combined data for S1 at $T_1$ and S2 at $T_2$. 
(n = 1959). The means for 35 year olds in the complete S1 was set to $t = 0$ in order to illustrate the magnitude of mean predicted memory losses (or gains) at different ages.

As can be seen, the cross-sectional data are consistent with a continuous age-related decrease for the episodic measure. Whereas a linear function provides a good approximation to these data, a weak accelerated trend is in fact discernible (cf., Verhaeghen & Salthouse, 1997). Significant decline (-2.58 t) is reached already by age 45. The point of a moderate effect size (0.5 $SD$; Cohen & Cohen, 1975) is in turn reached by age 55 and the predicted memory loss is substantial in magnitude (> 1 $SD$) already by age 65. In total, the predicted memory loss from 35 to 85 years exceeds 2 population $SD$s. The cross-sectional data on semantic memory indicate stable performance until 55, that is 10 years later than for episodic memory, and exhibit slightly less drop than episodic memory past that age, the total predicted loss at 85 amounting to about 16 t-units (1.6 $SD$).

The foregoing patterns should be directly compared with those depicted in Figure 2. These latter patterns were derived from summing the change scores across age groups (e.g., Schaie, 1994, 1996). In line with previous descriptions of the time-related changes, these data are consistent with a small continuous improvement in performance across the age period from 35 to 60, for both the episodic and semantic factor. The age of significant mean level deficit starting with the 35-year olds (-2.58 t-score units) is, in fact not reached until age 80 for episodic memory and age 85 for episodic memory. Decline is predicted past age 60 for both factors, more so for the episodic measure. For the oldest cohorts the decline is actually somewhat steeper than expected from cross-sectional data (See Figure 1). For both forms of declarative memory, the depicted pattern of changes are clearly curvilinear, consistent with course of memory change radically different from that depicted by the cross-sectional data.

Estimation of and Adjustment for Practice Effects
In order to validate the longitudinal patterns of memory changes, we estimated the magnitude of practice effects. As outlined previously, we first needed to know by how much the returnees in S1 differed from their age-matched equivalents in S2 who lacked prior exposure to the tests included. Thus, for each cohort and memory factor, the mean for S1 at T2 (Tables 3 and 4) was subtracted by the mean of S2 at T2 (Table 5). For example, for the youngest age cohort, on the episodic factor: S1T2 (returnees) – S2T2: (59.83 t) – (57.41 t) = 2.42 t. Apart from random/sampling error, this difference (D) may be assumed to result from the joint effects of mean level impact of selective attrition in the longitudinal sample (A) and practice (P), such that D = A + P.

In order to estimate practice effects (P), we first need to estimate the mean level impact of attrition upon T2 means in the longitudinal sample (A). This was done on basis of T1-data. Specifically, for each cohort and memory measure, the T1-mean for the complete S1 was subtracted from the T1-mean for the returnees (Tables 3 and 4). This is done under the assumption that dropout is unrelated to rate of change. As noted previously, the impact of selective attrition depends on (a) the magnitude of the retest status difference and (b) the extent of dropout. Because the rates of attrition were low across most of the cohorts, and as the attrition effect was modest in magnitude, the impact of attrition upon T2 means is small (but positive). For the youngest cohort, to provide an example: S1T1 (returnees) – S1T1 (whole group): (57.70 t) – (57.29 t) = 0.41 t. To obtain estimates of the practice effect, the estimated effect of attrition was finally subtracted from the mean sample difference (D – A = P): (2.42 t) – (0.41 t) = 2.01 t.

To minimize error due to random fluctuations in the sampling process, we averaged the values of D, A, and P separately for the five youngest age cohorts (i.e., 35-55 years initially) and the five oldest cohorts (60-80 years initially). The resulting estimates are presented in Figure 3 (episodic memory) and Figure 4 (semantic memory). In line with prior observations, returnees in S1 were superior to participants in S2, as judged from the positive D-values for the episodic and semantic measures. Moreover, the sample difference is larger for the older than for the
younger adults, both for the episodic and the semantic memory factor. Estimates of attrition (A) were low in general, but higher for the older cohorts as a result of the higher dropout frequencies with advancing age. Finally, and of primary concern for present purposes, the practice effect estimates (P) obtained by subtracting A from D are positive in all contrasts, amounting to about 1.5 \( t \)-scores for the episodic measure for both the younger and older age cohorts. By contrast, the estimates for the semantic factor were negligible (about 0.4 \( t \)).

Whereas an effect size of 1.5 \( t \)-scores may appear small, adjustment for practice effects alters the composite age gradients in a critical fashion (Figure 5) in the case of episodic memory. Specifically, the pattern of adjusted cumulative change scores reveals a period of stable performance from age 35 to age 60, rather than continuous increments. As a consequence of the fact that practice effects were task-specific, these adjusted trajectories exhibit a clear differential pattern of changes for the episodic and semantic memory measures, unlike the unadjusted longitudinal data portrayed in Figure 2.

Cross-Sequential Analyses of Independent Sample Data

As outlined in the introduction, we planned to perform cross-sequential analyses on the between-sample data. At this point, it is important to note that adjusting Time 2 means (for returnees), by cohort-wise subtraction of the absolute retest-effect estimates presented above, results in the same mean values as those observed for S2. This follows from the fact that any Time 2 difference (D) between samples is attributed to effects of practice plus attrition (i.e., assuming that dropout is unrelated to change). As we here contrast data for the full S1T₁ sample and the full S2T₂ sample we need not care about attrition effects. In other words, the pattern of changes derived from between-sample data are in principle the same as those obtained by adjusting mean change scores in the longitudinal sample for practice effects.

The results of the 10 (cohort) X 2 (Time) ANOVA on episodic memory revealed significant effects of Cohort, \( F(9, 1939) = 126.86, \, \eta^2 = .371 \), and of Time, \( F (1, 1939) = 26.42, \, \eta^2 = .013 \). Again, the Cohort X Time interaction was significant, \( F(9, 1939) = 3.72, \, \eta^2 = .017 \),
data indicating little time-related effect for the youngest cohorts, but decline for groups 60 years and above. The magnitude of the interaction effect was attenuated as compared with the within-samples analyses. The effect of time was furthermore magnified by a small margin as compared with the within-sample analyses, presumably due to a lack of impact of practice effects in the case of the independent-sample comparison. The corresponding analyses of semantic memory revealed effects of Cohort, $F(9, 1939) = 64.02, \eta^2 = .229$, but not of Time, $F(1, 1939) = 1.68, p > .10$. Finally, the Cohort X Time interaction was marginally significant, $F(9, 1939) = 2.24, p = .018, \eta^2 = .010$, in line with the trend of minor increment in across the younger half of the sample and decrements for the older participants described earlier. Taken together, the outcome of the cross-sequential ANOVA for the independent samples seem to be consistent with a slightly different pattern of changes as compared with the within-sample analyses for episodic memory, consistent with the difference between unadjusted (Figure 2) and practice-adjusted (Figure 5) composite longitudinal patterns discussed earlier.

**Influence of Cohort Differences in Education**

Even after adjusting for practice effects, the longitudinal patterns in Figure 5 deviate considerably from the cross-sectional patterns in Figure 1, with considerably later onset of decline for the episodic measures longitudinally, and with improvements, instead of a period of stability for semantic memory up to about age 55. In order to examine the possibility that this remaining discrepancy reflects the fact that cohort differences in educational attainment influence cross-sectional data, we conducted simple and hierarchical regression analyses with episodic and semantic memory measures as criteria and age and education as regressors.

For this purpose S1 and S2 were divided into two halves or cohort sequences: those born from 1909 to 1934 and those born from 1939 to 1954. This permitted examination of whether the pattern of absence and presence of deficits for the younger and older adults, evident from Figure 5, would emerge when educational differences were accounted for. Separate analyses were conducted for the two samples to cross-validate the results.
The results of these analyses are summarized in Table 6 (episodic memory) and Table 7 (semantic memory). The findings agree well with predictions based on the adjusted longitudinal data. First, in episodic memory (Table 6), the age-related variance remaining when age was entered after education in the younger half of the sample was only a fraction of that remaining in the older half of the sample (0.005/0.106 = .047 for sample 1 and 0.004/0.141 = 0.028). In fact, age did not add any significant variance for the youngest cohorts, suggesting that differences in episodic memory among the five younger cohorts may be attributable to differences in education rather than differences in chronological age. Second, for semantic memory (Table 7) the regression coefficients for age reverted from being negative to slightly positive when age was entered after education in the younger half of the sample (age added a significant amount of variance in Sample 1; in Sample 2 there was a tendency, \( p < .02 \), in the same direction), in line with the patterns in Figure 5. By contrast, the coefficients were negative in direction for the older half of the sample regardless of whether education was entered first or not. Together, these results indicate that cross-sectional data adjusted for differences in education agree well with practice-adjusted longitudinal data, supporting a pattern of late onset of decline of episodic memory performance and a pattern of minor improvements in middle age, with a subsequent and less pronounced deficit in semantic memory performance.

General Discussion

Five-year changes in episodic and semantic memory were examined in a population-based sample \( (n=829) \) of adults ranging in age from 35 to 80 years initially. A cohort-matched sample of participants \( (n = 959) \) was included at Time 2 to control for retest effects. The results raise a number of theoretically and methodologically important issues regarding adult memory development. We first provide a summary of differences between the predicted patterns of change emergent from cross-sectional and longitudinal data. Next, we discuss these patterns related to the following major issues: \( (a) \) practice effects, \( (b) \) cohort
differences in education, (c) differential aging patterns in episodic and semantic measures, and (d) patterns of memory changes in relation to changes in basic factors.

**Cross-sectional versus Longitudinal Patterns of Change**

On the basis of the present results, the answer to the question whether cross-sectional and longitudinal data diverge with regard to age changes in declarative memory depends on which age range one considers. From age 60 to 85, the predicted decline of episodic memory was at least 1.5 SD-units regardless of type of design used to estimate changes. In a similar vein, results demonstrate reliable decline in semantic memory performance across the equivalent age period (an effect that may approach 1 SD). Thus, both declarative memory factors appear to be vulnerable to the effects of aging past middle age and in the case of episodic memory the extent of functional loss is considerable.

For groups 60 to 70 years old, cross-sectional and longitudinal estimates of change were similar. For the oldest groups (75- and 80-year olds), there was some deviance, with somewhat more time-related decline than expected from cross-sectional data (see also, Zelinski & Burnight, 1997). This matter possibly reflects the fact that attrition was selective and most prominent (40%) for the oldest age groups. If the more able tended to decline more across the retest interval, such an effect should be expected to be effective in groups subjected to more extensive attrition. A median split analysis on Time-1 data confirmed that, across age, higher performing participants improved somewhat less or declined more than those in the below-median sample. Also, all participants willing to participate a second time were re-assessed, and, although effort was taken to exclude cases of dementia at retest, it is possible that a few of these cases did not meet the initial inclusion criteria for sensory functioning.

Whereas the cross-sectional and longitudinal age gradients were largely similar for the older cohorts, cross-sectional and longitudinal predictions differed considerably for the younger participants. Cross-sectional data indicated a deficit in episodic memory performance of almost 1 SD up to age 60, whereas the practice-adjusted longitudinal data demonstrated that
performance is stable up to this age. In a similar vein, longitudinal estimates of semantic memory change were consistent with improvements in performance, instead of stability, across the same age period. It is difficult to conceive of a period-factor, operating during the present time frame that would produce these age-selective discrepancies between cross-sectional and longitudinal data. Even though dropout turned out to be negatively selective, the magnitude was modest (about 0.4 SD) and invariant across age (cf., McDonald-Miszczak et al., 1995). Most importantly, attrition rates were very low (about 10%) across the young/middle-aged groups. Thus, attrition is not a plausible reason for the discrepancy.

Practice Effects and their Implications for Longitudinal Studies

The results indicated that the returnees benefit from previous exposure to the present battery of tests, at least on the episodic measures. The finding that practice effects may be present over an interval as long as five years is important. It appears likely that past longitudinal studies underestimated time-related deficits due to such effects. Rabbitt, Diggle, Smith, Holland, and McInnes (2001) recently arrived at a similar conclusion in longitudinal study of fluid intelligence, using a different method of estimating practice effects (based on the fact participants were tested and unequal number of times during the course of the study). The observation that the magnitude of the practice effects was age-invariant is in agreement with past observations (Schaie, 1988). Thus, practice effects may yield the impression that younger/middle-aged participants improve and minimize time-related decline for the old in longitudinal studies. However, the concern that practice effects may account for the basic difference in direction of the time-related change in younger and older adults, observed here and in a few other studies (e.g., Zelinski & Burnight, 1997), appears to be unwarranted.

It is important for future studies to examine the conditions under which practice effects occur. Are these effects task-specific or is there transfer to similar tasks? Do the effects reflect learning or do they reflect some non-specific factor such as alteration in motivation or acquaintance with testing? This is an important issue, as use of alternate test forms will not
eliminate the effects provided that transfer effects operate or if the practice-related influence is largely nonspecific. Whereas a best guess is that several factors sum to produce practice effects (e.g., Horn & Donaldsson, 1976), transfer may appear less likely provided that studies deliberately trying to produce memory gains (i.e., training studies) often obtain no or weak effects across shorter time intervals (see Kausler, 1994). No practice effects were observed for the semantic measure, suggesting that the effects are task (domain) specific. Presumably, retrieval attempts of highly over-learned material (as in semantic tasks) as opposed to novel/unique materials (as in episodic tasks) have a limited effect on future performance. If practice effects are nonspecific, one might further expect a retest advantage on a task not encountered before. Lack of advantage for returnees over new participants on a cognitive measure introduced at Time 2 (Tower of Hanoi; Rönnlund, Lövdén, & Nilsson, 2001) does not support this hypothesis. In sum, the results indicated that practice effects distort inferences regarding time-related changes and explained part of the discrepancy between the cross-sectional and longitudinal data. Further research is needed to understand the specific nature of these effects, but there are some indications that they are task specific.

The Role of Cohort Differences in Educational Attainment

The fact that longitudinal and cross-sectional estimates of change differed even when practice effects were accounted for suggested that cohort factors are important to consider. An extensive body of studies of children provides converging evidence that added exposure to education has a direct influence on performance on a variety of cognitive measures, including memory tasks (Cahan & Cohen, 1989; for a review, see Ceci, 1991). Specifically, prolonged education increases test sophistication, expands knowledge, and likely enhances perceptual and verbal skills that support basic encoding and retrieval processes. If these effects are persistent, and if each successive generation receives more average formal schooling, alterations in the educational system may serve to create a positive cohort gradient for measures of declarative memory (cf., Schaie, 1994).
Consistent with the hypothesis that secular increases in education attainment underlie most of the apparent early onset of episodic decline depicted cross-sectionally, age was no longer a predictor of performance among the younger cohorts when educational differences were partialed out. In a similar vein, adjustment for educational differences altered the cross-sectional aging patterns in semantic memory in a critical fashion. When differences in education were adjusted for, minor improvements up to about age 55 were predicted. This was in line with the longitudinal data, but unlike the raw cross-sectional data, where a plateau in performance was apparent across the same age period. Thus, when the influence of potential threats to the internal validity (differences in education and practice effects) were controlled for, the cross-sectional and longitudinal aging patterns converged. This held true for both forms of declarative memory.

The present indications that cohort-related differences in education may affect not only semantic memory but also episodic memory is noteworthy, as some investigators have considered episodic tasks to be immune to cohort differences (e.g., Kausler, 1991). As such, these result are in line with indications that secular trends in other fluid aspects of cognition, including reasoning tasks (e.g., Teasdale & Owen, 1989; Williams, 1998) may reflect added exposure to education. However, regardless of age, schooling was more strongly related to semantic than episodic memory, consistent with a relative difference with regard to the strength of influence of educational attainment upon episodic and semantic memory performance.

Recent cross-sectional data by Park et al. (2002) may appear to contradict one aspect of the present findings. In that study an early onset of episodic decline was namely apparent even though a lack of differences in educational background was reported. However, the latter assertion was based on category data, and, at least in Sweden, part of the secular increase in quantity of formal education occurred within educational levels (e.g., prolonged primary school, Jonsson & Mills, 1993). Thus, it could be that some differences in quantity of education actually existed among groups in the study by Park et al. This possibility as well as a
difference with regard to representativeness of samples across studies (participants in the study by Park et al. were highly select as judged from data on educational background, whereas the present study was population-based) may explain the apparent discrepancy. It should be noted that there are examples of cross-sectional studies involving education-matched samples that observed negligible differences in episodic memory before 60 (e.g., Albert, Heller, & Millberg, 1988; Giambra, Arenberg, Zonderman, Kawas, & Costa, 1995; Lehman & Mellinger, 1986), in line with the present results.

Whereas differences in educational attainment seem sufficient to explain age differences in episodic memory in the age range from 35 to 60 years, we cannot dismiss the possibility that other cohort factors such as improved nutrition (e.g., Martorell, 1998) or increases in environmental complexity (e.g., Schooler, 1998) contribute. Further knowledge of the cohort-related trends in memory and their relation to schooling and other factors may be gained from cohort-sequential analyses (Schaie, 1965, 1977) of the present and third-wave data of the Betula Study.

Differential Aging Patterns in Measures of Episodic and Semantic Memory

A differential aging pattern of episodic and semantic was evident from findings of (a) small, but reliable improvements from 35 to about 55 in semantic memory, a period during which episodic memory performance appears to remain stable, and (b) substantially more decline for the episodic memory measure past onset age. These observations hold for the adjusted cross-sectional and longitudinal data (cf., also Nyberg et al., 2001).

As noted, few longitudinal studies have compared changes in measures of episodic and semantic memory and some results are not particularly supportive of a differential aging pattern (e.g., Schaie, 1994). As viewed from the present results, a possible reason is that these studies did not take practice effects into account. Specifically, if unadjusted changes were consulted, composite change patterns were similar for episodic and semantic memory, presumably due to the pattern of task-specific practice effects discussed above. The issue of
whether the differential age aging pattern is best understood by considering that the two factors reflect differentially age sensitive structures (or processes) or that the factors load differentially on a common age sensitive factor (cf., Salthouse & Czaja, 2000; Baltes & Lindenberger, 1997) remains open. Conceivably, shared as well as unique factors contribute to age-related changes in episodic and semantic memory.

Patterns of Memory Change and Their Relation to Changes in Basic Factors

The adjusted cross-sectional and longitudinal data converged on showing that a decline in declarative memory functioning begins after age 60 (see also Schaie, 1994; Zelinski & Burnight, 1997). An interesting question, then, is why a decline emerges around this age. An age-related reduction of declarative memory has been linked to several basic factors, including biological/structural factors (Raz, 2000; Prull, Gabrieli, & Bunge, 2000), and processing resources such as speed and working memory capacity (Park et al., 2002). Further studies examining the relations between changes in such basic factors and changes in memory, and the relations between changes in cognitive resources and changes in biological factors, are needed to gain further knowledge of the causes of the memory decline that occurs in old age.

In the context of the finding of maintained memory functioning until age 60, it is intriguing to note that reductions in biological factors may be have an earlier onset (age 20 or 30) and adhere to a linear decline function. This may hold for brain weight (e.g., Greenfield et al., 1967) and integrity of the dopamine system (e.g., Bäckman & Farde, 2001; Bäckman et al., 2000; Eggers, Haug, & Fischer, 1984). An early onset and linear progression of decline has similarly been observed on measures of processing factors such as working memory and perceptual speed, both from point of cross-sectional (Park et al, 1996, 2002) and longitudinal data (speed, Schaie, 1994).

There are several possible explanations for the apparent discrepancy between the finding of a late onset of declarative-memory decline and findings of an early onset of decline for more basic factors. One obvious possibility is that a curve-linear function is a better account
of age changes also for more basic factors. There is some evidence to support this possibility (Antonini et al., 1993; Rinne, Lönnberg, & Marjamäki, 1990). A second possibility is that although there is an early onset decline for more basic factors, the negative influence of such reductions on declarative memory is counteracted by compensatory mechanisms. At the neural level there is functional (Cabeza et al., 1997; see Cabeza, 2001, 2002 for reviews) as well as structural (Bartzokis et al., 2001) data that can be interpreted as reflecting compensation. At the psychological level, a factor that could serve an important compensatory role is increased knowledge with advancing age (Salthouse, 1993; Bryan, Luszcz, & Crawford, 1997). A related possibility is that the decline in basic factors has to reach a certain threshold before it affects cognitive functions such as memory (cf., Park, 2002). The notion of a critical threshold is supported by studies of white-matter changes (Boone et al., 1992; DeCarli et al., 1995) and of dopamine depletion related to Parkinson's disease (see Forno, 1996).

Taken together, reductions in basic factors, both at level of processing resources and biological factors are likely important to consider in accounting for the decline in episodic and semantic memory that occurs in old age. There are several potential explanations for the finding that memory performance remains stable until age 60. These include the possibility of a curve-linear age function also for basic factors, as well as the possibility that the decline in basic factors must reach a certain threshold, and outweigh possible compensatory processes, before it has a manifest effect.

**Summary and Conclusions**

Longitudinal and cross-sectional data from large-scale representative samples revealed substantial decrements in performance in old age, especially in episodic memory. However, no episodic decline was apparent before age 60 and semantic memory tended to improve up to about this age. Cohort differences in educational attainment may to be a main determinant of the apparent early “deficits” depicted by cross-sectional data. In conclusion, the results provide strong evidence that the age-trajectories for episodic and semantic memory differ
and point to the need to control for cohort and practice effects, respectively, in cross-sectional and longitudinal studies. These observations raise questions that await forthcoming multilevel research efforts, but should provide important insights concerning the development of memory in adulthood as well as measurement of cognitive change in general.
References


Adult memory development


Figure captions

Figure 1. Predicted mean memory change (T-scores) in episodic and semantic memory across age based on cross-sectional data.

Figure 2. Predicted mean memory change (T-scores) in episodic and semantic memory across age based on longitudinal data.

Figure 3. Mean sample difference (D) and estimated mean effect of attrition (A) and practice (P) for episodic memory by age group.

Figure 4. Mean sample difference (D) and estimated mean effect of attrition (A) and practice (P) for semantic memory by age group.

Figure 5. Predicted mean memory change (T-scores) in episodic and semantic memory across age based on longitudinal data adjusted for practice effects.
Footnotes

1 As noted by Arenberg (1982), independent sampling of S1 and S2 at different points in time, as in the present study, and in accord with Schaie (1965), may be less than ideal. This is because the set of living persons within a given cohort may not be the same at different points of measurement, due to mortality. An alternative would be to sample all future participants at Time 1 and to randomly assign these to the time(s) of measurement at which they are first assessed. However, due to low mortality rates across most of the included age range, these effects should be small in this study. Most importantly, our cross-sample comparisons revealed that S1 and S2 were highly similar with regard to sample composition across the variables of interest.

2 It might be argued that Time 2 data should be entered for Sample 1 participants as well. Such data were collected. The means at Time 2 for Sample 1 differed only marginally from those at Time 1, as presented in Table 1 (as would be expected in a sample above age 35), and they are therefore not presented here. Most important, analyses of Time 2 data (with Time 1 data being imputed for dropouts) did not alter the results reported here.

3 It is equally valid, logically, to assume that there are no practice effects. In this case, any mean difference between Sample 1 and Sample 2 at Time 2 is taken to indicate differential rate of change in dropouts and returnees. Given the sizes of the sample differences obtained, and given the fact that dropout rates were about 10% across 7 of the cohorts, an average decline in dropouts of more than 2 SD-units across the five-year period is required to account for the sample difference at Time 2.

4 This finding may be an artifact regression of towards the mean rather than signaling a true ability x change interaction when, as in the present case, only two points of measurement is available.
Acknowledgements

This research was supported by a grant to Lars-Göran Nilsson from the Bank of Swedish Tercentenary Foundation, the Swedish Council for Research in the Humanities and Social Sciences, and the Swedish Council for Social Research. Lars Nyberg and Lars Bäckamn were supported by grants from the Swedish Science Council.
Table 1. Background Characteristics for Participants in S1 at T1 by Age Cohort and Retest-status.

<table>
<thead>
<tr>
<th>Age cohort</th>
<th>Returnees</th>
<th>Dropouts</th>
<th>Returnees</th>
<th>Dropouts</th>
<th>Returnees</th>
<th>Dropouts</th>
<th>Returnees</th>
<th>Dropouts</th>
</tr>
</thead>
<tbody>
<tr>
<td>35-40</td>
<td>46/46</td>
<td>4/4</td>
<td>14.07</td>
<td>12.31</td>
<td>28.66</td>
<td>28.00</td>
<td>35.58</td>
<td>20.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.55)</td>
<td>(2.84)</td>
<td>(1.18)</td>
<td>(1.77)</td>
<td>(8.60)</td>
<td>(7.39)</td>
</tr>
<tr>
<td>40-45</td>
<td>46/44</td>
<td>6/4</td>
<td>13.44</td>
<td>14.35</td>
<td>28.52</td>
<td>28.60</td>
<td>31.62</td>
<td>33.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.03)</td>
<td>(6.54)</td>
<td>(1.26)</td>
<td>(2.17)</td>
<td>(7.83)</td>
<td>(11.24)</td>
</tr>
<tr>
<td>45-50</td>
<td>46/42</td>
<td>8/4</td>
<td>12.93</td>
<td>11.29</td>
<td>28.27</td>
<td>28.00</td>
<td>32.61</td>
<td>27.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4.20)</td>
<td>(4.29)</td>
<td>(1.62)</td>
<td>(1.65)</td>
<td>(9.42)</td>
<td>(6.52)</td>
</tr>
<tr>
<td>50-55</td>
<td>55/35</td>
<td>5/5</td>
<td>10.37</td>
<td>10.90</td>
<td>28.22</td>
<td>27.80</td>
<td>30.20</td>
<td>27.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.76)</td>
<td>(3.53)</td>
<td>(1.32)</td>
<td>(1.69)</td>
<td>(8.18)</td>
<td>(11.37)</td>
</tr>
<tr>
<td>55-60</td>
<td>48/40</td>
<td>6/6</td>
<td>9.10</td>
<td>7.75</td>
<td>28.40</td>
<td>27.75</td>
<td>30.07</td>
<td>26.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.36)</td>
<td>(2.05)</td>
<td>(1.46)</td>
<td>(2.14)</td>
<td>(8.88)</td>
<td>(11.02)</td>
</tr>
<tr>
<td>60-65</td>
<td>42/49</td>
<td>4/5</td>
<td>8.73</td>
<td>9.78</td>
<td>28.15</td>
<td>27.56</td>
<td>25.82</td>
<td>24.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.14)</td>
<td>(3.73)</td>
<td>(1.54)</td>
<td>(1.88)</td>
<td>(7.96)</td>
<td>(6.29)</td>
</tr>
<tr>
<td>65-70</td>
<td>46/37</td>
<td>8/9</td>
<td>8.30</td>
<td>7.47</td>
<td>27.99</td>
<td>26.47</td>
<td>23.67</td>
<td>18.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.01)</td>
<td>(2.50)</td>
<td>(1.37)</td>
<td>(2.07)</td>
<td>(8.23)</td>
<td>(10.69)</td>
</tr>
<tr>
<td>70-75</td>
<td>39/37</td>
<td>9/15</td>
<td>8.20</td>
<td>8.27</td>
<td>27.64</td>
<td>27.33</td>
<td>23.34</td>
<td>19.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.43)</td>
<td>(2.86)</td>
<td>(1.72)</td>
<td>(1.31)</td>
<td>(9.02)</td>
<td>(7.78)</td>
</tr>
<tr>
<td>75-80</td>
<td>36/35</td>
<td>17/12</td>
<td>7.56</td>
<td>7.34</td>
<td>26.80</td>
<td>26.69</td>
<td>17.44</td>
<td>16.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.88)</td>
<td>(2.33)</td>
<td>(2.36)</td>
<td>(2.71)</td>
<td>(8.12)</td>
<td>(8.32)</td>
</tr>
<tr>
<td>80-85</td>
<td>38/22</td>
<td>21/19</td>
<td>7.70</td>
<td>6.86</td>
<td>26.87</td>
<td>25.25</td>
<td>15.92</td>
<td>11.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.19)</td>
<td>(2.95)</td>
<td>(2.24)</td>
<td>(2.36)</td>
<td>(7.31)</td>
<td>(7.28)</td>
</tr>
<tr>
<td>Total/M*</td>
<td>442/387</td>
<td>88/83</td>
<td>10.04</td>
<td>9.63</td>
<td>27.95</td>
<td>27.34</td>
<td>26.63</td>
<td>22.58</td>
</tr>
</tbody>
</table>

Note: Marginal mean
Table 2. Background Characteristics for Participants by Age cohort in Sample 2 at Second Time of Measurement. For the Cognitive Measures the Corresponding Values (M, SD) for Returnees in Sample 1 are provided in the Rightmost Columns.

<table>
<thead>
<tr>
<th>Age cohort</th>
<th>Sex distrib.</th>
<th>Education (Years)</th>
<th>MMSE (Max = 30)</th>
<th>Block Design (Max = 51)</th>
<th>PMC (1-5)</th>
<th>MMSE</th>
<th>Block Design</th>
<th>PMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>51/49</td>
<td>13.56</td>
<td>28.56</td>
<td>34.04</td>
<td>3.26</td>
<td>28.53</td>
<td>35.26</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.50)</td>
<td>(1.32)</td>
<td>(8.72)</td>
<td>(0.60)</td>
<td>(1.28)</td>
<td>(8.54)</td>
<td>(0.45)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.58)</td>
<td>(1.42)</td>
<td>(9.15)</td>
<td>(0.68)</td>
<td>(1.64)</td>
<td>(8.19)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>50</td>
<td>51/49</td>
<td>12.21</td>
<td>28.26</td>
<td>31.22</td>
<td>3.37</td>
<td>28.49</td>
<td>33.42</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.69)</td>
<td>(1.28)</td>
<td>(9.53)</td>
<td>(0.58)</td>
<td>(1.37)</td>
<td>(8.83)</td>
<td>(0.59)</td>
</tr>
<tr>
<td>55</td>
<td>53/47</td>
<td>10.36</td>
<td>28.11</td>
<td>29.46</td>
<td>3.51</td>
<td>28.07</td>
<td>29.62</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.33)</td>
<td>(1.70)</td>
<td>(9.81)</td>
<td>(0.69)</td>
<td>(1.61)</td>
<td>(8.47)</td>
<td>(0.53)</td>
</tr>
<tr>
<td>60</td>
<td>56/44</td>
<td>9.75</td>
<td>27.89</td>
<td>26.37</td>
<td>3.55</td>
<td>28.09</td>
<td>28.09</td>
<td>3.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.91)</td>
<td>(1.51)</td>
<td>(8.11)</td>
<td>(0.56)</td>
<td>(1.48)</td>
<td>(8.42)</td>
<td>(0.53)</td>
</tr>
<tr>
<td>65</td>
<td>58/41</td>
<td>8.11</td>
<td>27.48</td>
<td>23.29</td>
<td>3.54</td>
<td>27.66</td>
<td>24.21</td>
<td>3.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.56)</td>
<td>(1.77)</td>
<td>(8.60)</td>
<td>(0.54)</td>
<td>(1.59)</td>
<td>(7.52)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>70</td>
<td>63/36</td>
<td>7.27</td>
<td>26.99</td>
<td>20.36</td>
<td>3.57</td>
<td>27.42</td>
<td>20.67</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.28)</td>
<td>(2.14)</td>
<td>(8.03)</td>
<td>(0.57)</td>
<td>(1.68)</td>
<td>(9.36)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>75</td>
<td>63/37</td>
<td>7.39</td>
<td>26.57</td>
<td>18.15</td>
<td>3.64</td>
<td>26.91</td>
<td>20.79</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.46)</td>
<td>(1.94)</td>
<td>(7.31)</td>
<td>(0.56)</td>
<td>(2.28)</td>
<td>(9.01)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>80</td>
<td>60/36</td>
<td>7.21</td>
<td>26.14</td>
<td>15.19</td>
<td>3.71</td>
<td>26.38</td>
<td>15.56</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.59)</td>
<td>(2.24)</td>
<td>(7.34)</td>
<td>(0.54)</td>
<td>(2.62)</td>
<td>(8.97)</td>
<td>(0.72)</td>
</tr>
<tr>
<td>85</td>
<td>43/22</td>
<td>6.66</td>
<td>25.15</td>
<td>12.40</td>
<td>3.74</td>
<td>26.15</td>
<td>13.97</td>
<td>3.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.34)</td>
<td>(2.51)</td>
<td>(6.42)</td>
<td>(0.62)</td>
<td>(2.27)</td>
<td>(8.58)</td>
<td>(0.70)</td>
</tr>
<tr>
<td>Total/M</td>
<td>550/409</td>
<td>9.61</td>
<td>27.35</td>
<td>24.16</td>
<td>3.52</td>
<td>27.58</td>
<td>25.41</td>
<td>3.46</td>
</tr>
</tbody>
</table>

Note. PMC = Perceived memory change; 1 = much better, 5 = much worse as compared with five years ago.
<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Action recall (max = 16)</th>
<th>Recall of verbal nouns (max = 16)</th>
<th>Recall of action nouns (max = 16)</th>
<th>Recognition of verbal nouns (max = 8)</th>
<th>Recall of new facts (max = 20)</th>
<th>Episodic factor (T-score)</th>
<th>Total S</th>
<th>Dropouts</th>
<th>Returnees</th>
<th>Mean change</th>
</tr>
</thead>
<tbody>
<tr>
<td>35-40</td>
<td>10.89</td>
<td>10.59</td>
<td>11.28</td>
<td>11.72</td>
<td>7.95</td>
<td>8.64</td>
<td>5.40</td>
<td>5.83</td>
<td>7.36</td>
<td>8.50</td>
</tr>
<tr>
<td></td>
<td>(2.56)</td>
<td>(2.47)</td>
<td>(2.37)</td>
<td>(2.03)</td>
<td>(3.01)</td>
<td>(3.11)</td>
<td>(2.01)</td>
<td>(1.65)</td>
<td>(3.45)</td>
<td>(3.74)</td>
</tr>
<tr>
<td>40-45</td>
<td>10.26</td>
<td>10.64</td>
<td>11.04</td>
<td>11.38</td>
<td>7.17</td>
<td>8.18</td>
<td>5.66</td>
<td>5.39</td>
<td>7.11</td>
<td>7.83</td>
</tr>
<tr>
<td></td>
<td>(2.48)</td>
<td>(2.33)</td>
<td>(2.17)</td>
<td>(2.22)</td>
<td>(3.30)</td>
<td>(2.92)</td>
<td>(1.62)</td>
<td>(1.96)</td>
<td>(2.88)</td>
<td>(3.26)</td>
</tr>
<tr>
<td>45-50</td>
<td>10.34</td>
<td>10.16</td>
<td>11.14</td>
<td>10.98</td>
<td>6.77</td>
<td>7.28</td>
<td>5.10</td>
<td>5.22</td>
<td>6.32</td>
<td>6.93</td>
</tr>
<tr>
<td></td>
<td>(2.24)</td>
<td>(2.56)</td>
<td>(1.89)</td>
<td>(2.50)</td>
<td>(2.81)</td>
<td>(3.31)</td>
<td>(1.81)</td>
<td>(2.47)</td>
<td>(3.05)</td>
<td>(2.98)</td>
</tr>
<tr>
<td>50-55</td>
<td>9.74</td>
<td>9.77</td>
<td>10.60</td>
<td>10.97</td>
<td>7.13</td>
<td>7.54</td>
<td>5.32</td>
<td>5.28</td>
<td>5.91</td>
<td>6.67</td>
</tr>
<tr>
<td></td>
<td>(2.42)</td>
<td>(2.38)</td>
<td>(2.38)</td>
<td>(2.08)</td>
<td>(3.30)</td>
<td>(2.82)</td>
<td>(1.97)</td>
<td>(2.11)</td>
<td>(2.91)</td>
<td>(3.00)</td>
</tr>
<tr>
<td>55-60</td>
<td>9.39</td>
<td>9.02</td>
<td>10.07</td>
<td>10.40</td>
<td>6.26</td>
<td>6.68</td>
<td>4.89</td>
<td>5.16</td>
<td>5.69</td>
<td>6.07</td>
</tr>
<tr>
<td></td>
<td>(2.11)</td>
<td>(2.51)</td>
<td>(2.36)</td>
<td>(2.60)</td>
<td>(3.33)</td>
<td>(3.20)</td>
<td>(1.88)</td>
<td>(2.07)</td>
<td>(2.75)</td>
<td>(2.94)</td>
</tr>
<tr>
<td>60-65</td>
<td>8.41</td>
<td>8.00</td>
<td>9.82</td>
<td>9.46</td>
<td>5.68</td>
<td>5.25</td>
<td>4.93</td>
<td>4.75</td>
<td>5.57</td>
<td>5.56</td>
</tr>
<tr>
<td></td>
<td>(2.64)</td>
<td>(2.61)</td>
<td>(2.14)</td>
<td>(2.22)</td>
<td>(2.64)</td>
<td>(3.04)</td>
<td>(1.77)</td>
<td>(2.14)</td>
<td>(2.95)</td>
<td>(2.71)</td>
</tr>
<tr>
<td>65-70</td>
<td>7.84</td>
<td>7.59</td>
<td>9.53</td>
<td>9.08</td>
<td>5.80</td>
<td>5.40</td>
<td>4.84</td>
<td>4.71</td>
<td>5.23</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td>(2.40)</td>
<td>(2.56)</td>
<td>(2.43)</td>
<td>(2.95)</td>
<td>(2.88)</td>
<td>(3.39)</td>
<td>(1.94)</td>
<td>(2.39)</td>
<td>(2.97)</td>
<td>(2.93)</td>
</tr>
<tr>
<td>70-75</td>
<td>7.05</td>
<td>6.28</td>
<td>9.14</td>
<td>8.45</td>
<td>5.29</td>
<td>4.72</td>
<td>4.63</td>
<td>4.13</td>
<td>5.04</td>
<td>4.91</td>
</tr>
<tr>
<td></td>
<td>(2.87)</td>
<td>(3.23)</td>
<td>(2.40)</td>
<td>(2.81)</td>
<td>(3.06)</td>
<td>(2.79)</td>
<td>(1.69)</td>
<td>(2.04)</td>
<td>(2.78)</td>
<td>(2.78)</td>
</tr>
<tr>
<td>75-80</td>
<td>6.34</td>
<td>5.46</td>
<td>8.32</td>
<td>7.23</td>
<td>4.06</td>
<td>3.61</td>
<td>3.75</td>
<td>3.56</td>
<td>4.46</td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td>(2.55)</td>
<td>(2.78)</td>
<td>(2.46)</td>
<td>(2.80)</td>
<td>(2.70)</td>
<td>(2.61)</td>
<td>(1.90)</td>
<td>(2.10)</td>
<td>(2.30)</td>
<td>(2.49)</td>
</tr>
<tr>
<td>80-85</td>
<td>5.90</td>
<td>4.45</td>
<td>8.48</td>
<td>6.30</td>
<td>4.61</td>
<td>3.03</td>
<td>4.15</td>
<td>3.00</td>
<td>3.97</td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td>(2.60)</td>
<td>(2.70)</td>
<td>(2.71)</td>
<td>(3.33)</td>
<td>(3.03)</td>
<td>(2.75)</td>
<td>(1.90)</td>
<td>(2.19)</td>
<td>(2.28)</td>
<td>(2.84)</td>
</tr>
</tbody>
</table>

Marginal M 8.62 8.20 9.94 8.90 6.07 6.03 4.87 4.70 5.66 5.73 50.00 46.84 50.79 50.14 -0.65

Note. S = Sample, \( T_1 \) = First time of measurement, \( T_2 \) = Second time of measurement.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Vocabulary (max = 30)</th>
<th>General knowledge (max = 10)</th>
<th>Word fluency (M, five letters)</th>
<th>Word fluency (Professions, B)</th>
<th>Semantic factor (T-scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35-40</td>
<td>23.47 (23.84)</td>
<td>7.93 (1.65)</td>
<td>12.84 (4.89)</td>
<td>14.11 (5.99)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>40-45</td>
<td>23.14 (23.84)</td>
<td>8.16 (1.60)</td>
<td>12.53 (4.89)</td>
<td>13.16 (5.99)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>45-50</td>
<td>23.06 (1.69)</td>
<td>8.23 (3.89)</td>
<td>12.50 (4.93)</td>
<td>13.16 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>50-55</td>
<td>23.08 (1.69)</td>
<td>8.39 (3.89)</td>
<td>12.50 (4.93)</td>
<td>13.16 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>55-60</td>
<td>22.53 (1.69)</td>
<td>8.24 (3.89)</td>
<td>12.43 (4.93)</td>
<td>13.16 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>60-65</td>
<td>21.76 (1.69)</td>
<td>8.10 (3.89)</td>
<td>12.43 (4.93)</td>
<td>13.16 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>65-70</td>
<td>21.04 (1.69)</td>
<td>7.51 (3.89)</td>
<td>10.66 (4.93)</td>
<td>11.20 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>70-75</td>
<td>20.24 (1.69)</td>
<td>6.81 (3.89)</td>
<td>10.66 (4.93)</td>
<td>11.20 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>75-80</td>
<td>19.44 (1.69)</td>
<td>6.15 (3.89)</td>
<td>9.04 (4.93)</td>
<td>9.77 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>80-85</td>
<td>18.65 (1.69)</td>
<td>5.49 (3.89)</td>
<td>7.73 (4.93)</td>
<td>7.41 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
<tr>
<td>85-90</td>
<td>17.86 (1.69)</td>
<td>4.84 (3.89)</td>
<td>6.20 (4.93)</td>
<td>5.71 (4.93)</td>
<td>4.31 (3.13)</td>
</tr>
</tbody>
</table>

Note. T1 = First time of measurement. T2 = Second time of measurement.
<table>
<thead>
<tr>
<th>Age</th>
<th>Action recall (max =16)</th>
<th>Recall of action nouns (max =16)</th>
<th>Recall of verbal nouns (max =8)</th>
<th>Recall of recent facts (max =20)</th>
<th>Episodic Factor (T-scores)</th>
<th>Vocab (max =30)</th>
<th>General knowledge (max=10)</th>
<th>Word fluency (M-5)</th>
<th>Word fluency (P-B)</th>
<th>Semantic Factor (T-scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>10.22</td>
<td>11.15</td>
<td>7.98</td>
<td>5.85</td>
<td>7.26</td>
<td>57.41</td>
<td>7.79</td>
<td>12.79</td>
<td>6.20</td>
<td>5.15</td>
</tr>
<tr>
<td></td>
<td>(2.25)</td>
<td>(2.27)</td>
<td>(2.77)</td>
<td>(1.70)</td>
<td>(2.94)</td>
<td>(7.06)</td>
<td>(2.96)</td>
<td>(1.42)</td>
<td>(4.26)</td>
<td>(2.71)</td>
</tr>
<tr>
<td>45</td>
<td>10.01</td>
<td>11.14</td>
<td>7.13</td>
<td>4.98</td>
<td>6.48</td>
<td>54.89</td>
<td>7.95</td>
<td>12.88</td>
<td>6.22</td>
<td>5.25</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(2.47)</td>
<td>(3.31)</td>
<td>(2.00)</td>
<td>(3.03)</td>
<td>(8.78)</td>
<td>(4.27)</td>
<td>(1.58)</td>
<td>(5.01)</td>
<td>(2.78)</td>
</tr>
<tr>
<td>50</td>
<td>9.82</td>
<td>11.18</td>
<td>7.45</td>
<td>5.36</td>
<td>6.71</td>
<td>54.64</td>
<td>24.69</td>
<td>8.55</td>
<td>12.31</td>
<td>6.16</td>
</tr>
<tr>
<td></td>
<td>(2.31)</td>
<td>(1.77)</td>
<td>(2.80)</td>
<td>(1.71)</td>
<td>(2.74)</td>
<td>(6.15)</td>
<td>(2.97)</td>
<td>(1.42)</td>
<td>(4.78)</td>
<td>(3.47)</td>
</tr>
<tr>
<td>55</td>
<td>9.43</td>
<td>10.79</td>
<td>7.09</td>
<td>5.44</td>
<td>6.67</td>
<td>54.66</td>
<td>23.30</td>
<td>8.08</td>
<td>11.91</td>
<td>6.17</td>
</tr>
<tr>
<td></td>
<td>(2.37)</td>
<td>(2.22)</td>
<td>(3.00)</td>
<td>(1.73)</td>
<td>(3.23)</td>
<td>(8.31)</td>
<td>(3.85)</td>
<td>(1.46)</td>
<td>(5.06)</td>
<td>(3.18)</td>
</tr>
<tr>
<td>60</td>
<td>8.43</td>
<td>10.23</td>
<td>6.22</td>
<td>5.06</td>
<td>5.98</td>
<td>51.49</td>
<td>22.40</td>
<td>8.07</td>
<td>11.98</td>
<td>6.02</td>
</tr>
<tr>
<td></td>
<td>(2.73)</td>
<td>(2.54)</td>
<td>(2.94)</td>
<td>(1.98)</td>
<td>(3.13)</td>
<td>(8.81)</td>
<td>(5.27)</td>
<td>(1.60)</td>
<td>(4.72)</td>
<td>(3.13)</td>
</tr>
<tr>
<td>65</td>
<td>7.45</td>
<td>9.08</td>
<td>5.10</td>
<td>4.17</td>
<td>5.05</td>
<td>46.79</td>
<td>21.74</td>
<td>7.68</td>
<td>11.02</td>
<td>5.82</td>
</tr>
<tr>
<td></td>
<td>(2.60)</td>
<td>(2.53)</td>
<td>(3.22)</td>
<td>(2.04)</td>
<td>(2.84)</td>
<td>(9.04)</td>
<td>(5.21)</td>
<td>(1.65)</td>
<td>(4.83)</td>
<td>(3.16)</td>
</tr>
<tr>
<td>70</td>
<td>6.75</td>
<td>8.42</td>
<td>4.82</td>
<td>3.96</td>
<td>4.59</td>
<td>44.65</td>
<td>19.32</td>
<td>7.10</td>
<td>9.60</td>
<td>4.74</td>
</tr>
<tr>
<td></td>
<td>(2.66)</td>
<td>(2.60)</td>
<td>(2.90)</td>
<td>(1.98)</td>
<td>(2.73)</td>
<td>(9.14)</td>
<td>(6.19)</td>
<td>(1.85)</td>
<td>(4.64)</td>
<td>(2.92)</td>
</tr>
<tr>
<td>75</td>
<td>5.90</td>
<td>7.65</td>
<td>3.97</td>
<td>3.94</td>
<td>3.99</td>
<td>41.92</td>
<td>18.92</td>
<td>7.05</td>
<td>8.94</td>
<td>4.19</td>
</tr>
<tr>
<td></td>
<td>(2.63)</td>
<td>(2.87)</td>
<td>(2.57)</td>
<td>(2.21)</td>
<td>(2.30)</td>
<td>(8.80)</td>
<td>(5.85)</td>
<td>(1.79)</td>
<td>(3.90)</td>
<td>(2.49)</td>
</tr>
<tr>
<td>80</td>
<td>4.95</td>
<td>6.72</td>
<td>3.50</td>
<td>3.39</td>
<td>3.36</td>
<td>38.61</td>
<td>17.71</td>
<td>6.17</td>
<td>8.56</td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td>(2.91)</td>
<td>(2.71)</td>
<td>(2.52)</td>
<td>(2.26)</td>
<td>(2.21)</td>
<td>(8.83)</td>
<td>(6.41)</td>
<td>(2.20)</td>
<td>(4.35)</td>
<td>(2.90)</td>
</tr>
<tr>
<td>85</td>
<td>3.72</td>
<td>5.86</td>
<td>2.46</td>
<td>2.60</td>
<td>2.86</td>
<td>34.50</td>
<td>14.95</td>
<td>5.82</td>
<td>6.55</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>(2.43)</td>
<td>(2.68)</td>
<td>(1.91)</td>
<td>(2.13)</td>
<td>(1.84)</td>
<td>(7.20)</td>
<td>(6.21)</td>
<td>(1.96)</td>
<td>(4.49)</td>
<td>(2.46)</td>
</tr>
<tr>
<td>Mean</td>
<td>7.67</td>
<td>9.22</td>
<td>5.57</td>
<td>4.48</td>
<td>5.30</td>
<td>47.96</td>
<td>21.10</td>
<td>7.43</td>
<td>10.65</td>
<td>5.29</td>
</tr>
</tbody>
</table>

Note. Vocab = Vocabulary, M-5 = M and five letters, P-B = Professions, initial letter B.
Table 6. Summary of Simple and Hierarchical Regression of Age and Years of education on Episodic Memory by Cohort-sequence and Sample.

<table>
<thead>
<tr>
<th>Cohort-sequence</th>
<th>Sample 1</th>
<th>Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>$C_1$-$C_5$ Age</td>
<td>-.246</td>
<td>.059</td>
</tr>
<tr>
<td>Edu</td>
<td>.331*</td>
<td>.138</td>
</tr>
<tr>
<td>Age</td>
<td>-.096</td>
<td>.142</td>
</tr>
<tr>
<td>$C_5$-$C_{10}$ Age</td>
<td>-.375</td>
<td>.146</td>
</tr>
<tr>
<td>Edu</td>
<td>.265*</td>
<td>.100</td>
</tr>
<tr>
<td>Age</td>
<td>-.433*</td>
<td>.206</td>
</tr>
</tbody>
</table>

*p < .01
Table 7. Summary of Simple and Hierarchical Regression of Age and Years of Education on Semantic Memory by Cohort-sequence and Sample

<table>
<thead>
<tr>
<th>Cohort-sequence</th>
<th>Sample 1</th>
<th>Sample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample 1</td>
<td>Sample 2</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>$C_1$-$C_5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-.061</td>
<td>.002</td>
</tr>
<tr>
<td>Edu</td>
<td>.516*</td>
<td>.186</td>
</tr>
<tr>
<td>Age</td>
<td>.177*</td>
<td>.208</td>
</tr>
<tr>
<td>$C_6$-$C_{10}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-.305</td>
<td>.091</td>
</tr>
<tr>
<td>Edu</td>
<td>.433*</td>
<td>.215</td>
</tr>
<tr>
<td>Age</td>
<td>-.228*</td>
<td>.265</td>
</tr>
</tbody>
</table>

*p < .01
Episodic Factor

Prediction Mean Memory Change (T Scores)

\[ y = -21.739 + 1.1003x - 1.4903 \times 10^{-2}x^2 + 3.2510 \times 10^{-5}x^3 \]

\[ R^2 = 0.993 \]

Semantic Factor

\[ y = -0.94020 + 0.16256x - 3.0743 \times 10^{-3}x^2 - 2.1327 \times 10^{-5}x^3 \]

\[ R^2 = 0.997 \]
Figure 2

Predicted Mean Memory Change (T Scores)

Ep: \( y = 16.388 - 1.6277x + 4.5881e^{-2}x^2 - 3.5747e^{-4}x^3 \)  \( R^2 = 0.992 \)

Sem: \( y = -27.682 + 1.1230x - 9.4960e^{-3}x^2 - 2.5175e^{-6}x^3 \)  \( R^2 = 0.980 \)
Figure 4

Age Group

Estimated Mean Effect (T Scores)

Younger

Older

D
A
P
Figure 5

Predicted Mean Memory Change (T Scores) vs. Age

Episodic Factor: $y = 11.516 - 0.95497x + 2.5950e^{-2}x^2 - 2.2552e^{-4}x^3$  \( R^2 = 0.997 \)

Semantic Factor: $y = -22.212 + 1.1233x - 0.9079e^{-3}x^2 - 2.3776e^{-6}x^3$  \( R^2 = 0.985 \)