

Supervisor: Dr. Roger A. Dixon

**"The Way We Were":  
Retrospective Evaluations of Memory Performance  
in the Elderly**

by

**Carolyn Barnes Crow  
B.A., University of Calgary, 1988**

**A Thesis submitted in Partial Fulfilment of the  
Requirements for the Degree of**

**MASTER OF ARTS**

**in the Department of Psychology**

**We accept this thesis as conforming  
to the required standard**

[Redacted Signature]

**Dr. R.A. Dixon, Supervisor, (Department of Psychology)**

[Redacted Signature]

**Dr. D.F. Hultsch, Departmental Member, (Department of Psychology)**

[Redacted Signature]

**Dr. M. Penning, Outside Member, (Department of Sociology)**

[Redacted Signature]

**Dr. P.J. Moss, External Examiner, (Department of Geography)**

[Redacted Signature]

**© CAROLYN BARNES CROW, 1998**

**Dr. M. Penning, Outside Member, (Department of Sociology)**  
**University of Victoria**

**All rights reserved. This thesis may not be reproduced in whole or in part, by  
photocopy or other means, without the permission of the author.**

**Dr. P.J. Moss, External Examiner, (Department of Geography)**

using six nouns from each of five taxonomic categories. Category similarities within lists were avoided to prevent joint testing of categories such as vegetables and fruits. Word order within lists was randomized, with the constraint that no two words from the same category could be placed side-by-side. Participants were given 2 minutes to study each list and 5 minutes to write the items they recalled. The average number of words remembered across both lists was used as the measure of word recall.

Fact Recall. Participants were asked to answer two sets of general information questions on the first two testing occasions, in a group setting. The sets each consisted of 40 questions that tested individuals' knowledge of different domains. These questions were taken from a total of 240 questions in the domains of science, history, sports, geography and entertainment (Nelson & Narens, 1980). The average number of correct items across the two question sets was used as the fact recall measure.

Metamemory. The Metamemory in Adulthood (MIA) instrument (Dixon et al., 1988) was administered during the first testing session, in a group setting. The 108-item version of the MIA was used, which consists of seven subscales: Capacity (memory self-efficacy), Change (perceived change in memory across time), Task (knowledge of memory processes and functions), Strategy (use of mnemonics and external memory aids in everyday life), Anxiety (degree of anxiety involving memory), Achievement (achievement motivation regarding memory), and Locus (perceived control over memory). The MIA uses a Likert scale format. It asks participants either for their opinions about statements related to memory (1= *disagree*

ABSTRACT

Memory aging beliefs are investigated in the present study. A longitudinal sample of older adult volunteers from a medium-sized metropolitan area was tested on word, story and fact recall 4 times over a 9-year period. On the fourth occasion, they were asked to estimate their present and past performance levels and to rate their current performance relative to: (a) hypothetical comparative targets, and (b) their own past performances. Results indicated that: (a) participants believed their memory performance had declined, (b) task performance was systematically overestimated, (c) performance estimation accuracy was greater for current than for past occasions for word and fact recall, but not for story recall, (d) global memory self-efficacy was only moderately related to a few memory beliefs variables, and (e) current memory performance beliefs combined with memory change beliefs to predict past performances for both word and story recall, but not for fact recall. Results are interpreted in the framework of Ross' (1989) implicit theory hypothesis of memory for attributes of the self.

Examiners:

[Redacted]

Dr. R.A. Dixon, Supervisor, (Department of Psychology)

[Redacted]

Dr. D.F. Hultsch, Departmental Member, (Department of Psychology)

[Redacted]

Dr. M. Penning, Outside Member, (Department of Sociology)

[Redacted]

Dr. P.J. Moss, External Examiner, (Department of Geography)

	<u>Page</u>
ABSTRACT . . . . .	ii
TABLE OF CONTENTS . . . . .	iii
LIST OF TABLES . . . . .	vi
LIST OF FIGURES . . . . .	vii
ACKNOWLEDGEMENTS . . . . .	viii
DEDICATION . . . . .	ix
INTRODUCTION . . . . .	1
REVIEW OF THE LITERATURE . . . . .	
Metamemory . . . . .	3
Memory Self-Efficacy . . . . .	3
MSE Research . . . . .	4
Metamemory in Adulthood Instrument . . . . .	5
Predictions/Postdictions . . . . .	6
Implicit Theories . . . . .	8
Research Relating to Implicit Theories for Memory Performance . . . . .	10
Research on Cognitive Retrospections . . . . .	13
The Present Study . . . . .	15
PRINCIPAL RESEARCH QUESTIONS AND HYPOTHESES . . . . .	19
METHOD . . . . .	
Procedure . . . . .	21
Participants . . . . .	22
Materials . . . . .	25
Personal Data Sheet . . . . .	25
Story Recall . . . . .	25
Word Recall . . . . .	25
Fact Recall . . . . .	26
Metamemory . . . . .	26
Retrospective Evaluations . . . . .	28

	<u>Page</u>
Numerical Postdictions .....	28
Targeted Comparisons .....	29
<b>RESULTS</b>	
Overview .....	30
Profile of Memory Beliefs .....	31
Retrospective Evaluations .....	31
Numerical Postdictions .....	33
Targeted Comparisons .....	35
The Accuracy of Numerical Postdictions .....	39
Occasion Differences in Accuracy as Assessed by Correlations .....	39
Evaluating Accuracy by Difference Scores .....	42
Comparing Postdicted with Actual Scores .....	43
The Association of Memory Beliefs with MSE .....	45
Partial Correlations of Memory Belief variables with MSE .....	45
Age, Occasion and MSE Differences in Postdictions .....	47
Predicting Past Postdiction using Current Postdiction and Change .....	48
Differences Between Subjects in Selected and Unselected Samples .....	49
<b>DISCUSSION</b>	
Profile of Memory Beliefs .....	55
Accuracy of Memory Assessments .....	60
The Association of Memory Beliefs with MSE .....	66
Predicting Past Postdiction using Current Postdiction and Change .....	68
Sample Differences .....	71
Importance of the Present Study .....	74
Conclusion .....	76
REFERENCES .....	79
TABLES .....	84
FIGURES .....	102

# TABLE OF CONTENTS

Page

## APPENDICES

A	The 17 Items of the MIA Capacity Subscale . . . . .	113
B	The 18 Items of the MIA Change Subscale . . . . .	114
C	Wording of the Retrospective Evaluation . . . . .	116
D	Wording of the Numerical Postdiction Questionnaire . . . . .	118
E	Wording of the Targeted Comparison Questionnaire . . . . .	119

<u>Number</u>	<u>Title</u>	<u>Page</u>
1	Comparison of Educational and Occupational Characteristics of the VLS Sample (Age 55-86) with the General Population of British Columbia and Canada in 1986 . . . . .	84
2	Characteristics of Cohort 1 of the VLS Sample, 1986 - 1995 . . . . .	85
3	Retrospective Evaluations' Means, Standard Deviations and T-Test Values of Means Compared to Value Signifying Status is "About the Same" as Present . . . . .	86
4	Means and Standard Deviations for Numerical Postdictions, expressed in both Absolute and Percent Values . . . . .	87
5	Relative Postdiction Means, Standard Deviations and T-Test Values of Means Compared to Value Signifying Status is "About the Same" as Comparative Target . . . . .	88
6	Significant Task x Target Contrasts for Targeted Comparisons . . . . .	90
7	Pearson Product Moment Correlations of Numerical Postdictions with Actual Performances and the Significance of the Differences Between Past and Present Accuracy . . . . .	91
8	Percent Overestimation for Numerical Postdictions . . . . .	93
9	Age, Occasion and MSE Differences in Postdictions as Revealed in Hierarchical Multiple Regressions . . . . .	94
10	The Prediction of Past Numerical Postdiction using Current Numerical Postdiction, MIA Capacity and MIA Change . . . . .	95
11	Means, Standard Deviations and t-tests for Mean Differences between Selected and Unselected Samples . . . . .	97
12	Breakdown of Educational and Marital Differences Between Selected and Unselected Samples . . . . .	101

<u>Number</u>	<u>Title</u>	<u>Page</u>
1	Task x Age Group Interaction for Retrospective Evaluations . . . . .	102
2	Task x Occasion Interaction for Retrospective Evaluations . . . . .	103
3	Task x Gender Interaction for Numerical Postdictions . . . . .	104
4	Task x Gender Interaction for Targeted Comparisons . . . . .	105
5	Task x Target Interaction for Targeted Comparisons . . . . .	106
6	Occasion Effect of Accuracy . . . . .	107
7	Postdicted and Actual Performance Scores (as Percentages) Over 4 Occasions for Word Recall . . . . .	108
8	Postdicted and Actual Performance Scores (as Percentages) Over 4 Occasions for Story Recall . . . . .	109
9	Postdicted and Actual Performance Scores (as Percentages) Over 4 Occasions for Fact Recall . . . . .	110
10	Sample Differences in Numerical Postdictions Across Occasions . . . . .	111
11	Sample x Occasion Interaction in Performance Levels . . . . .	112

## ACKNOWLEDGEMENTS

viii

This research was supported by a grant from the U.S. National Institute on Aging (AG 08235) to Roger Dixon.

I would like to thank my husband, Malcolm, for his endless support, both emotional and instrumental. Thanks for believing in and encouraging me. Thanks for producing my graphs over and over again. Thanks for looking after Meaghan so I could work and for supporting us so I could pursue my degree.

Heartfelt thanks also to Dr. Roger Dixon for seeing me through my Master's degree with such expertise, humanity and humour. You put the super in supervisor!

To my baby girl, Meaghan Alexandra Crow, born 2 days before my originally scheduled defense date; 2 months before the date it actually happened. Although you couldn't wait to be born, you let me finish my degree once you arrived! You are the baby every expectant mother dreams of.

## Introduction

Cognitive aging, the changes undergone in the thinking processes with age, is an area of increasing interest to researchers. Until recently, it was a somewhat neglected area of developmental research, with much interest being focused on the earlier portion of the lifespan. Research in developmental psychology largely centred on the period from birth through adolescence, with the vast adult portion of the lifespan being treated as a fairly stable phase, and the latter portion of adulthood treated as a period of decline. Recently, researchers focusing on the cognitive changes experienced throughout adulthood are discovering that aging is not solely characterized by decline, but also by maintenance and, in some cases, growth (Baltes, 1987). Nevertheless, generally held beliefs about certain areas of functioning continue to focus upon decline with aging.

Memory is one area for which young and old adults alike perceive loss with increasing age (Heckhausen, Dixon & Baltes, 1989). This decline-oriented belief is particularly important for older individuals. For example, such beliefs have the potential to affect whether old people feel able to cope with the demands of life, and to evoke a great deal of anxiety over anticipated or retrospected decline. In addition, the legitimacy of these beliefs affects their everyday lives both through the potentially serious problems encountered as a result of memory loss (e.g., forgetting to take medication), and through perceived or actual memory loss leading to the adoption of compensatory mechanisms (e.g., following a strict routine to ensure no necessary activity is neglected). The particular importance of older adults' age-decline beliefs warrants attention to their details, accuracy and manifestations.

One area in which age-decline beliefs may be manifested is the self-report of memory changes. Beliefs about change may be determined more by an individual's implicit theories of memory aging than by an awareness of actual change (McDonald-Miszczak, Hertzog & Hultsch, 1995). Implicit theories of memory aging may even be combined with ideas about memory's current status to form ideas about memory's past status (Ross, 1989). If this is the case, memory change complaints are likely driven by beliefs more than by the accurate monitoring of change over time. Clinicians attempting to evaluate memory change would be well served to be aware of this. Unfortunately, there is very little research to date which addresses the impact of implicit theories of memory decline on such things as reports of memory loss, or actual changes in memory.

The present study will attempt to shed some light on this area. It measures the memory performance of community dwelling older adults living in a medium-sized metropolitan area on 4 occasions over a 9-year period. On the most recent occasion of measurement, an extended set of memory belief variables was also assessed. In general, the study is intended to increase the overall understanding of memory beliefs in the elderly. The objectives are to assess: (a) the profile of memory beliefs in the elderly, (b) the accuracy of these beliefs, (c) the influences of such variables as age and memory self-efficacy on memory beliefs, (d) the process whereby past memory performances are estimated, and (e) differences between participants with complete or nearly complete metamemory data and those missing responses.

## Review of the Literature

In this section, literature will be reviewed in two principal areas pertaining to the objectives outlined above. Metamemory will be first examined, followed by a review of information pertaining to implicit theories regarding memory.

### Metamemory

Metamemory, or cognitions about memory (Hertzog & Dixon, 1994), encompasses the knowledge and ideas individuals have about memory's capacity, the way in which it functions, and is developed and used by themselves and others (Dixon & Hultsch, 1983). The significance of metamemory can be conceptualized in terms of its implications for adaptive behaviour (Hertzog & Dixon, 1994). For example, it is possible that awareness plays a role in the extent to which compensatory behaviours are engaged in as an adaptation to perceived decline in memory capacity.

Additionally, suboptimal metamemory knowledge and use could contribute to inefficient memory strategies being adopted by older adults. Brigham and Pressley (1988) demonstrated that older adults failed to recognize a particular mnemonic device as more effective than an alternative device, even when their performance improved as a result of its use. Even the segment of the sample who did understand the technique's superiority seemed to avoid its use in favour of a more familiar or easily executed strategy. Such findings suggest that metamemory may be mediating the relationship between memory and everyday functioning in older adults, which implies that it is a logical area to target for modification with training (McDougall, 1994).

Memory Self-Efficacy. A vital component of metamemory (Hertzog & Dixon, 1994)

is memory self-efficacy (MSE) or the belief in one's ability to use memory effectively in demanding situations. Self-efficacy in general is influenced by personal factors, such as past experiences and affect, and social factors, such as societal stereotypes surrounding abilities (Cavanaugh, 1996). Compared with a low sense, a high sense of self-efficacy for an area is associated with choosing more challenging tasks, setting higher goals and persevering longer with tasks (Bandura, 1989). Consistent with Berry and West's (1993) overview of the self-efficacy literature, it follows that older adults holding positive beliefs regarding their memory's capacity would be advantaged compared to their peers with low MSE, because they might be more inclined to invoke instrumental behaviours designed to produce their own memory outcomes. One would predict that they would be likely to feel positively about memory challenges, to set their sights high for performance, and to keep trying longer than would older adults with low MSE. These ideas have prompted investigations into whether MSE level can predict memory performance level.

MSE Research. While some studies have demonstrated a relationship between MSE and memory performance (e.g. Hertzog, Dixon, & Hultsch, 1990b), a simple model of memory beliefs predicting memory performance does not emerge. In other words, people's beliefs about their memory capabilities are not necessarily accurate (Hertzog & Dixon, 1994).

One typical method of ascertaining the accuracy of memory self-efficacy beliefs is to examine the correlations of memory task performance with memory self-efficacy scales (e.g., Hertzog, Saylor, Fleece, & Dixon, 1994). Hertzog and Dixon

(1994) report the range of these correlations as being between 0 and 0.3, both in normal and memory-impaired adults. The low to moderate size of these correlations indicates that general MSE is not the only factor involved in memory performance. However, it is recognized that this relationship could be moderated as a result of factors like the type of memory task, or the specificity of the self-efficacy measure to the task.

The task specificity of MSE measures is an area of both theoretical and practical importance. As pointed out by Berry and West (1993), there is a lack of consensus as to the level of specificity at which self-efficacy should be measured. General MSE has, for example, been tapped using conceptually broad metamemory questionnaires with relevant subscales, such as the Metamemory in Adulthood instrument (MIA; Dixon & Hultsch, 1983), while task-specific MSE has often been measured using ratings of perceived capability to perform specific memory tasks, such as are found in the Memory Self-Efficacy Questionnaire (MSEQ; Berry, West, & Dennehy, 1989; as cited in Berry & West, 1993). As observed by Berry & West, none of the measures are inherently better than the others, but their choice should be dictated by the level of analysis adopted. Domain-level measures are valuable in providing general MSE information while task-specific measures can provide more precise information on efficacy-performance relationships.

Metamemory in Adulthood Instrument. The MIA was developed expressly to study memory self-perceptions among older adults. It is composed of items that emphasize everyday, ecologically relevant metamemory activities. The instrument was explicitly

designed to reflect the multidimensional nature of metamemory, by including questions relevant to memory knowledge, memory-related affect, and memory-related self-efficacy. Factor analyses have led to the identification of seven reliable and factorially valid subscales (Strategy, Task, Change, Anxiety, Achievement, Capacity and Locus). The Capacity subscale is also highly correlated with the Change subscale (Dixon, 1989).

Predictions/Postdictions. Some interesting research has tapped knowledge of and beliefs concerning memory for specific tasks using participants' estimates of task performance that are either predicted (taken prior to memory performance), or postdicted (taken after memory performance). They can be considered to be a complex amalgam of self-referent information. Hertzog and Dixon (1994) suggest that memory performance predictions may in fact contain elements of knowledge about memory and memory performance, self-efficacy, and on-line awareness of memory's contents and operations (monitoring). According to this perspective, an appraisal of task demands is made, which is combined with self-efficacy judgments to produce performance predictions. Performance postdictions are theoretically very similar to predictions and appear to provide a more precise measure of performance monitoring. For example, Hertzog, Saylor, Fleece, and Dixon (1994) found that postdicted performance estimates for word recall across both old and young participants were more accurate than estimates which were predicted. This finding supports both the idea that tasks were more accurately appraised as a result of experience with them, and that memory monitoring took place during task performance. However, since both

predictions and postdictions were made by the same participants, it is not clear whether the memory monitoring evident in the postdictions took place spontaneously, or was prompted by the experimentally-required predictions. Additionally, the postdictions may have been contaminated by a desire by participants to appear consistent with their predictions. The findings of Devolder, Brigham and Pressley (1990) suggest that the performance awareness upgrade evident in postdictions was likely due to performance monitoring that occurred without experimental prompting and was not biased by a desire for consistency. In a series of three experiments, they tested older and younger adults on nine different recall, recognition and prospective memory tasks. Participants were assigned to either a predicted or a postdicted performance estimation condition. Older adults generating postdictions were more accurate in their performance estimates than those generating predictions in five out of the nine tasks, thus supporting the ideas that participants were spontaneously monitoring memory content and that responses were not contaminated by the desire for consistency between predictions and postdictions. In addition, consistent with other research (e.g., Hertzog et al., 1994) younger and older adults were found to be equally accurate in their postdicted estimates.

The combined evidence surrounding the use of postdictions suggests that they are a useful measure of memory monitoring. Studies employing postdicted performance estimates (e.g., Brigham & Pressley, 1988; Devolder et al., 1990; Devolder, 1993) have tended to employ a cross-sectional design examining age differences between older and younger adults. Studies employing numerical

postdictions in a longitudinal design are distinctly lacking in the literature.

### Implicit Theories

Related to the area of metamemory are the theories people hold regarding their memory. Implicit theories are "schemalike knowledge structures that include specific beliefs regarding the inherent stability of an attribute, as well as a set of general principles concerning the conditions likely to promote personal change or stability" (Ross, 1989, p. 342). The literature suggests that a theory of memory decline with age operates among young and old alike (e.g. Heckhausen et al., 1989).

Implicit theories are believed by theorists such as Ross (1989) to have an important role in very long-term memory. This perspective views very long-term memory as reconstructive rather than reproductive. Bahrick (1984) argues that overlearned memory content such as the answers to simple arithmetic problems and the meanings of common words are reproduced literally. In contrast, he sees memory for emotionally and motivationally-laden content and autobiographical information as undergoing reconstructive changes. That is, memory of autobiographical, emotionally or motivationally-laden material from the past does not come in a photograph-like format, correct in every detail.

According to Ross (1989), the process of remembering self-referent information begins when an individual notes his or her current status on an attribute. This is likely to be more salient and easier to access than past status. Next, the individual may invoke an implicit theory which includes beliefs about the inherent stability of the attribute, and some principles regarding the conditions that could lead to stability or

change. The combination of the present as a benchmark, and the implicit theory regarding the attribute's trajectory, form the estimate of the past status. Information retrieved that is ambiguous may be seen as supporting the theory. According to this hypothesis, when remembering our past appearance, we do not elicit a completely accurate self-image and compare it to our present reflection. Rather, we look at our reflection, and use our beliefs about how we have physically aged to produce a reconstructed image of our past self. As we age, we may infer that in the past we had much fuller hair, a slimmer waist and far fewer wrinkles. Whether or not this was actually the case, we may believe it to be so based on our current appearance and our implicit theories of physical aging.

What elements would be involved in a theory of change in memory across the lifespan? Implicit theories of memory aging can focus on consistency with or difference from the present. A consistency principle is invoked when memory ability in the past is seen as being the same as in the present. According to Ross (1989), this principle is adhered to when the current self is seen as being the same as the past self, thus fostering a temporally consistent self-image. A difference principle is adhered to when a theory of change in the self is adopted. The self is not seen as temporally consistent, but it is nevertheless coherent, as any change experienced is easily explained by a strongly held belief.

Very long-term recollections of the past self result from: (a) the past status being remembered clearly, (b) the past status being forgotten, and estimated at random, or (c) the past status being forgotten, and estimated systematically. Ross

(1989) cites various forms of systematic bias which would indicate that implicit theories are used to guide recall: People may exaggerate their memory's consistency over time and infer that their past memory performance was the same as it is currently. Alternately, people may overestimate the change that has taken place, which could result from either: (a) change in memory performance being uniformly miscalculated, or (b) stability in memory performance being incorrectly perceived as change. If past memory performance is remembered clearly, then to the extent that performance was assessed accurately at the time, performance estimates will be accurate. If past memory performance is forgotten and is estimated without a pattern, estimation errors should be random. If past memory performance is forgotten, and an implicit theory of consistency is adopted, past and present performance estimates will be the same. If past performance is forgotten and a theory of difference is adhered to, past and present performance estimates will differ. The accuracy or level and direction of distortion of past estimates will depend on the degree of similarity between past and present performances, and its correspondence with the theory adopted.

Research Relating to Implicit Theories for Memory Performance. Support for the idea that memory beliefs can bias the perception of memory performance over time comes in part from two longitudinal studies tracking changes in memory performance and its perception (Taylor, Miller, & Tinklenberg, 1992; Zelinski, Gilewski, & Schaie, 1993). Both studies failed to find a correlation between longitudinal change in self-reported frequency of forgetting and actual memory performance change. Additional evidence

that people hold biased beliefs about memory decline comes from a cross-sectional study comparing (a) a group of older individuals' theories of change of certain characteristics with age, (b) another group of older individuals' self-perceptions of their present status on these characteristics, together with their self-retrospections about their status on these characteristics when they were younger, and (c) a group of younger participants' ideas about their current status on these characteristics (McFarland, Ross, & Giltrow, 1992). Using the younger group as a representation of the older groups' previous status on the various characteristics, memory emerged as an attribute for which the elderly appeared to exaggerate decline. The findings from these studies suggest that the perception of memory change is influenced by an implicit theory of memory decline with age rather than an awareness of actual changes in memory ability (McDonald-Miszczak et al., 1995).

Another study which generated some support for an implicit theory interpretation of memory recall examined longitudinal data for 2 samples (McDonald-Miszczak et al., 1995). The first sample, measured over a 2-year period, was aged 22-78 years during the first wave of data collection. The second sample consisted of older adults (aged 55 to 86 on the first testing occasion) measured over a 6-year period. Measures of memory performance and memory beliefs were taken on each occasion of testing for both samples, and their relationships assessed in parallel analyses. Longitudinal changes in MSE (as measured by the MIA Capacity subscale) were not significantly related to longitudinal changes in memory performance. Such a result was highly unlikely if individuals had monitored memory as it changed, and had

adjusted their memory beliefs accordingly. In addition, participants' ratings of retrospective memory change (as measured by the MIA Change subscale on the last occasion) were not correlated with their longitudinal MSE changes (ie., their ideas of how their memory capabilities had changed since the past did not match the differences from present to past in their concurrent beliefs about memory capacity). However, MSE changes were correlated with longitudinal perceived change (ie., differences from present to past in their concurrent beliefs about memory capacity matched the differences over time in their ideas of how they had changed since the past). This pattern was seen as consistent with (a) stable implicit theories over time, (b) stable beliefs over time about the self-as-rememberer, and (c) concurrently constructed self-ratings of change being driven by any longitudinal changes in perceived ability levels. The findings were consistent with the hypothesis that current perceptions of ability level were driving current perceptions of change. This study also revealed evidence that did not support an implicit theory hypothesis. Significant, yet weak, correlations were observed between longitudinal changes in memory performance and retrospective perceptions of memory change. This suggested that memory changes may be monitored to some extent.

Additional evidence for the monitoring of memory performance change was provided by Johansson, Allen-Burge and Zarit (1997). They tested a group of very old ( $M = 86.85$ , range = 84 to 90 at time 1) people in Sweden on 5 memory tests over 3 occasions, 2 years apart. They found that the retrospective report of decline was associated with actual decline on 3 of the 5 tasks over a 2-year interval. It is notable

that since very general measures of memory decline were used, the association depicted was crude and could not address the level of accuracy of retrospective assessments. However, the findings suggest that elderly people are aware of actual memory change, and, in reflecting on past performances, may be influenced by the awareness of this actual decline. This does not, however, exclude the possibility that people are also influenced by expectations of inevitable decline.

### Research on Cognitive Retrospections

An area bearing similarity to past memory performance awareness is past cognitive performance awareness. In order to examine accuracy and biases in the recall of cognitive performance, longitudinal performance data must be compared to retrospective performance estimates. The literature is distinctly lacking in studies examining longitudinal cognitive retrospections for their accuracy. A notable exception is Schaie, Willis, and O'Hanlon's (1994) examination of actual and perceived change over time in intellectual performance in older ( $M = 77.24$  years, range = 71-95 years), middle-aged ( $M = 60.63$  years, range = 50-70 years) and young ( $M = 41.76$  years, range = 29-49 years) adults over a seven-year period. As part of an ongoing cognitive testing program, participants took the Thurstones' Primary Mental Abilities test (PMA; Thurstone & Thurstone, 1949) which assessed performance for five factors of intelligence: Verbal Meaning, Spatial Orientation, Inductive Reasoning, Number and Word Fluency. They were then asked to compare their present performance (1984) with that in the past (1977). The majority of participants across the age groups believed that their abilities remained the same over time. If the elderly

are susceptible to believing in an exaggerated age-decline stereotype, one would expect their estimates of their abilities' trajectories to be falsely pessimistic. Age effects for the perception and experience of decline revealed age differences in the false perception of decline in two of the five abilities tested. More older than younger participants overestimated the decline that they experienced in Verbal Meaning. Inductive Reasoning showed a similar pattern, with a greater proportion of the elderly being overly pessimistic than either the middle-aged or the young. Interestingly, for Number ability, a greater proportion of the younger and middle-aged adults perceived false decline than did the older participants. This mixed pattern of results does not support the idea that the elderly subscribe to universal and overblown pessimism regarding their cognitive abilities. It implies that the both the ability to estimate past performance accurately and the beliefs held regarding the trajectories of abilities may be specific to the abilities in question.

Memory ability is one area for which a strong argument can be made that the elderly might perceive false decline. Memory is an area of particular salience and concern for old people, much more so than for their youngers. In the past, it has been found that the elderly perceive that they have lower memory capacity, less control over their memory, and more change in their memory than do the young (Dixon, Hultsch, & Hertzog, 1988). In addition, adults of all ages tend to believe that we become more forgetful from our mid 50s to our late 80s (Heckhausen et al., 1989), and that decline is just as likely for themselves as for others (Ryan & Kwong See, 1993). Thus, it would seem probable that memory is an ability for which exaggerated

implicit theories of age-decline are held by the elderly. It is also possible that the young elderly may exhibit more pronounced exaggeration than the old elderly, as for them, an age-decline implicit theory of memory is likely to be less a reflection of actual decline than it is for their elders. The idea that the very old experience greater memory decline than the young old is supported by the findings of Hultsch and his colleagues (Hultsch, Hertzog, Small, McDonald-Miszczak, & Dixon, 1992), who found greater longitudinal decline in an earlier-born cohort than in a later-born one in word, story and fact recall and verbal working memory.

### The Present Study

How can we tell whether our memories about ourselves are accurate? For the most part, these memories are unverifiable. It would, however, be most intriguing to step back in time with our notions of "the way we think we were," and compare these with "the way we actually were." A past photograph compared with our present reflection would enable us to determine our level of memory accuracy for our physical appearance, and to make inferences regarding the implicit theories that we have adopted. This approach is precisely the one that is taken in the present research with regard to memory abilities in the elderly. Past and present measures of memory capabilities in the elderly will be compared with their perceptions of these attributes' past and present levels. From the information available, the accuracy of beliefs about past and present memory abilities will be assessed, and inferences made about the theories that are operating with regard to the trajectory of memory in old age.

The ideas held regarding consistency with and difference from the past are of

interest in the current research with regard to memory performance. Specifically, it is expected that older adults will be aware of some decline in their memory capabilities. In addition, they are expected to have internalized society's age-decline stereotype for memory. The combination of these factors is expected to result in an overestimation of memory decline with age.

The present study utilizes postdicted performance estimates measured not only immediately following task completion, but retrospectively at delays of three, six, and nine years. Episodic memory was measured using story recall and word recall, and semantic memory was indexed by fact recall. Each of these tasks was administered twice during each wave of testing. During the fourth wave, performance estimates were obtained for current and prior task performance immediately following the second administration of each task. Although performance was likely monitored at each time of testing -- three, six and nine years prior to performance estimation -- it is expected that this information would be ambiguous and not directly available to participants at the time of performance estimation. Rather, an implicit theory hypothesis predicts that the retrospective postdictions employed in the present study reflect: (a) participants' notions of their current status for performance of the memory task in question, and (b) their beliefs surrounding their trajectory for performance of the memory task over time. It is hypothesized that participants will use age-decline memory beliefs to guide their estimates which will result in inflated perceptions of past memory performance. In addition, it is hypothesized that accuracy will be greater for more recent occasions.

Participants' implicit beliefs about performance in previous waves of the study were obtained in the form of "retrospective evaluations." These were general estimates of the trajectory of performance in each task over time. Additional specific information regarding the implicit theories employed by the sample, was obtained using "targeted comparisons." These were performance comparisons with hypothetical selves and others (average person same age, average 20-year-old, self at age 20, and self's best with practice or training). They are expected to reveal valuable insights into participants' beliefs about the stability and plasticity of personal and general memory trajectories.

General memory self-efficacy is measured in the current study using the MIA Capacity subscale. The construct of MSE is expected to be related to the implicit theories of memory adopted by the current sample as measured through the study's indices. Specifically, those participants with high MSE would be expected to immediately and retrospectively postdict higher memory performance levels than those with low MSE. Participants with a high level of MSE are likely to infer generally higher performance levels across all occasions for the tasks investigated in the retrospective evaluations. Additionally, high MSE participants would likely see themselves favourably when compared with hypothetical selves and others.

General memory decline beliefs were measured using the Change scale of the MIA. An investigation will be conducted to determine whether these general beliefs were used in combination with the current performance estimate to produce estimates of performance in the past. It is expected that MIA Change will combine with the

numerical postdiction for the present occasion to predict performance estimates for previous occasions.

In addition, sample self-selection will be investigated. A portion of the participants did not provide complete data for various metamemory variables. Participants providing too little information are not included in the analyses. Their exclusion may bias the results because it is possible that these participants may differ on various demographic and performance indices from those who responded more completely. Typically, such participants score lower on performance measures and demographic indicators than do those participants who provide complete data. Should this be so, the study's generalizability may be slightly more limited than would otherwise be the case.

### Principal Research Questions and Hypotheses

The following principal research questions are addressed through the present research.

1. What is the profile of memory beliefs in the elderly?

Hypothesis: The elderly will believe their memory has declined, both in specific tasks and in general.

2. Are postdictions of memory performance accurate?
  - 2a. Are there occasion differences in accuracy (as assessed by correlations)?

Hypothesis: Postdicted memory estimation will be less accurate the more distant the postdiction, and more accurate the more recent the postdiction.

- 2b. Is there systematic bias (a direction to inaccuracy) evident in memory

postdictions?

Hypothesis: Past memory performance will be overestimated by the elderly.

3. Are age and occasion associated with the accuracy of memory postdictions (as measured by differences between postdicted and actual performances)?

Hypothesis 1: As age increases, postdictions will become more accurate.

Hypothesis 2: Accuracy will be greater for more recent occasions.

4. What are the effects of memory self-efficacy, age, gender, health, education and marital status on memory beliefs? This is an exploratory question with no specific hypotheses. Gender and education are included because both are known to be related to cognitive performance in adults, and they may be related to metamemory beliefs.

- 4a. Is memory self-efficacy associated with memory beliefs?

Hypothesis: Higher current memory self-efficacy will be associated with the perception of memory maintenance, and lower memory self-efficacy will be associated with the perception of decline in the elderly.

- 4b. Are there age, occasion and memory self-efficacy differences in postdictions?

Hypothesis: Higher memory self-efficacy will be associated with higher postdictions across occasions and ages.

5. How are retrospective postdictions generated?

Hypothesis: Retrospective postdictions will be significantly related to the use of the current performance combined with decline beliefs.

6. Are there differences between participants who provided complete or nearly complete metamemory information and those with numerous missing data points?

Hypothesis: Participants missing multiple data points are likely to score higher on demographic and performance variables as compared to participants with complete metamemory information.

## Method

### Procedure

The data used in the present study were generated by the Victoria Longitudinal Study (VLS). The VLS has been conducting cognitive aging research on an ongoing basis since 1986. The initial sample of participants has been tested every 3 years since 1986, the fourth wave of data having been collected in 1995 (see Hultsch et al., in press, for details concerning the VLS method). Over the years, additional test materials have been added to the original battery, producing a rich data set including a core of consistently tested cognitive dimensions. In 1995, participants were tested over four sessions of approximately two hours each, with each participant's four sessions falling within a one-month period wherever possible. A battery of pencil-and-paper tests was administered in small group settings (approximately two to eight people) over the first two testing sessions. These group testing sessions have been administered by the same research assistant, a female, middle-aged former nurse, since the study's inception in 1986. The third and fourth testing sessions consisted of both pencil-and-paper and computer-response tasks. Each participant's third and fourth testing sessions were administered one-on-one by a paid research assistant (usually a psychology graduate student), though not necessarily the same assistant for both sessions. The MIA and personal data sheet were administered during session one. The story recall, word recall and fact recall tasks were each administered twice (once each in the group settings: sessions one and two). Numerical performance postdictions and targeted comparisons for each memory task were collected

immediately following their second administrations (i.e. in session 2). Participants received pairs of stories, word lists and factual questions based on a counterbalancing procedure designed for three waves of testing. Upon the fourth wave of data collection (1995), initial pairs of task assignments were reintroduced. Retrospective evaluations were collected in the fourth wave of data collection (1995) during the fourth testing session.

### Participants

Advertisements in television and newspapers, plus local community organizations in a medium-sized metropolitan area provided the population from which individuals were initially recruited. Potential participants were told that the research was being conducted in order to describe the changes we experience in our memory as we age, and to develop an understanding of individual differences in these patterns. Participants ranged in age from 55 to 86 during the initial testing phase. Requirements for study participation included (a) good health, without a recent illness such as a stroke or heart attack, (b) reasonable eye-sight (defined as the ability to read print the size found in newspapers), (c) reasonable hearing (the ability to hear normal conversation), and (d) the ability to produce written responses. A comparison of the initial sample with the general populations of British Columbia and Canada aged 55 and over is included in Table 1. The data indicate that in comparison to both these populations, the VLS sample had been more highly educated and more likely to hold a professional or skilled job. Participants were paid a small honorarium in return for their participation. Note that an additional similar-sized cohort (cohort 2) was later

recruited under the same participation constraints. This cohort has been tested in 1992 and 1995, but will not be examined in the present research. Table 2 indicates some basic characteristics of cohort 1 as it has evolved over the years, including the reasons for loss of participants. As each new phase of testing approached, contact was made by a letter, followed by telephone calls in order to set up appointment times. One designated representative initiated the invitations and scheduled the testing sessions. If a participant was not reachable at the address or telephone number on record, effort was made to locate him/her through such means as the telephone book, telephone directory assistance, and word of mouth. Effort was made to keep in contact with and maintain current addresses for participants approximately annually through a holiday newsletter in December. In addition, a reception and informative talk regarding some general results of the research was held for the volunteers in September of 1992.

A total of 176 participants returned for the fourth wave of testing in 1995. Three individuals (2 women and 1 man) did not complete all of the testing sessions, which reduced the sample size to 173. The 9-year longitudinal sample had (a) an average of 14.40 years of education ( $SD = 3.08$ ); (b) good self-rated health compared to a perfect state ( $M = .91$   $SD = .87$ , on a scale ranging from 0 = very good health to 4 = very poor health); and (c) vocabulary ability scores of  $M = 44.05$  ( $SD = 6.64$ ) out of a total of 54 items (ETS Kit; Ekstrom, French, Harman, & Dermen, 1976). An initial viewing of the data revealed that participants sometimes failed to answer questions relating to postdictions and retrospective evaluations.

Final Sample Determination. The selected sample was determined on the basis

of the completeness of the data available. It was found that many participants had failed to answer questions directly relevant to the present study. Only 90 participants had complete data for all relevant variables. Since this number represented only 52% of the wave 4 sample, it was deemed desirable to develop a more lenient criterion for the development of a selected sample. It was decided that performance data must be complete in order to generate a basis from participants could answer the relevant "memory beliefs" (retrospective evaluations, targeted comparisons and numerical postdictions) questions. Allowance was made for a single missing data point within any memory beliefs category of data within each task. This was done in order to achieve a reasonable balance between number of subjects and the amount of missing data. Missing values were then replaced using the sample mean for each variable as the imputed value.

Small, Dixon, Hultsch and Hertzog (1998) found that women outperformed men on indices of word recall and story recall. One goal of the present study was to examine whether there was evidence for differential performance estimation as a function of gender or age/cohort group membership. Intermittently throughout the analyses, gender and age group were used as between subjects factors. The gender breakdown of the sample was 58 women and 48 men. The sample was divided into two age groups using the median (age 75) and the mean (age 75.2) as guidelines for a Young-Old/Old-Old split. The Young-Old were those participants up to and including age 75 ( $n = 58$ ), while the Old-Old included those participants age 76 and over ( $n = 48$ ). Age and gender combined to produce the following cell sizes: Young-Old

women = 37, Young-Old Men = 21, Old-Old women = 21, Old-Old men = 27.

### Materials

Personal Data Sheet. Participants completed a questionnaire assessing background variables such as education, family information, professional history, ethnicity, living arrangements, change in activity levels, health, sensory and medication information.

Story Recall. Participants were asked to recall the contents of two different stories on the first two testing sessions. The two stories administered were selected from a possible six, all of which were constructed and analyzed according to a hierarchical propositional system developed by Kintsch (1974) (see Dixon, Hultsch, & Hertzog, 1989). Stories had between 154 and 174 propositions. Participants were given three minutes to read and seven minutes to recall each story. Scoring of the stories was based on gist recall, with a proposition being scored as present if the gist of its meaning was expressed correctly. The average percent of propositions recalled in the two stories was used as the score. Trained coders scored each protocol for the memory of specific propositions. Interscorer reliability was calculated using percent agreement of individual propositions recalled in a selection of two protocols per story generated by participants in the first wave of data collection. The overall agreement across the scorers in all four waves of testing was high (ie. .86).

Word Recall. Participants were asked to recall the contents of two different word lists during the first two testing sessions. Lists were developed from the Howard (1980) and Battig and Montague (1969) norms. Six 30-word lists were constructed,

*strongly* to 5 = *agree strongly*) or how often they perform certain memory-related activities (1 = *never* to 5 = *always*). Scores were scaled as a mean Likert response per item, with higher scores indicating greater memory capacity, less decline over time, more task knowledge, more strategy use, a higher level of anxiety, higher achievement motivation, and a more internal locus of control respectively. Internal consistency (Cronbach's Alpha) for the seven subscales has been reported between .71 for Locus and .93 for Change (see Dixon et al.). Confirmatory factor analyses conducted by Hertzog, Hultsch and Dixon (1989) have indicated that the Capacity and Change scales are also indicators of a higher-order dimension of MSE. However, due to the nature of the questions in the current research, it is necessary to separate current MSE and beliefs about change over time. Therefore, the Capacity subscale will be used as the sole measure of MSE. The complete MIA Capacity subscale can be found in Appendix A, while the Change subscale can be found in Appendix B. Convergent validity between the MIA and the Memory Functioning Questionnaire (MFQ) has been found by Hertzog et al. on a memory self-efficacy factor, found to be best indicated by the MIA Capacity scale and the MFQ Frequency of Forgetting scale. These analyses also indicated the convergence of the MIA and MFQ as measures of perceived change in memory and strategy use. Discriminant validity has also been demonstrated for the MIA (Hertzog, Dixon & Hultsch, 1990a). Hultsch and his colleagues (as cited in Dixon et al., 1988) reported that the constructs indicated in the MIA are not accounted for by state or trait anxiety, generalized locus of control, depression or concurrent depressive affect. Predictive validity has been demonstrated between MIA scales and

performance on various cognitive tasks, such as verbal comprehension, induction, and memory span (Dixon, Hertzog & Hultsch, 1986), and story recall (Dixon & Hultsch, 1983).

Retrospective Evaluations. During the fourth wave of testing (1995), a retrospective questionnaire was added to the battery of tests. It was administered during each participant's last testing session. This questionnaire was designed to assess implicit beliefs about performance in previous waves on the following cognitive tasks: vocabulary, word recall, story recall, fact recall, word opposites, verbal working memory, lexical decision, comprehension speed, reading comprehension, and letter sets. The present study incorporates only the questions pertaining to word recall, story recall and fact recall. For each task evaluated, a sample question was provided to refresh the participant's memory. The participant was then asked to indicate how well he/she performed on this task in the past compared to the present. This question was presented three times for each task, indicating the time lapse which corresponded with each wave of testing in the past. For complete task wording, see Appendix C. A five-point Likert scale (ranging from 0 = *my performance was much better 3/6/9 years ago* to 4 = *my performance was much worse 3/6/9 years ago*) was used.

Numerical Postdictions. During the fourth wave of data collection (1995), a series of postdicted performance assessments were collected. Numerical postdictions were collected for word recall, story recall, and fact recall. Participants generated the postdictions directly following the completion of each task during the second testing session, in a group setting. Specifically, participants were asked to evaluate their

current performance as well as their performance on these tasks at lags of 3, 6, and 9 years. Participants were instructed not to refer back to their responses on the tasks when generating performance estimates, which were written on a dedicated page immediately following these task responses. Numerical performance estimates were used as the scores. Specifically, for word recall, an estimate of the number of words remembered out of 30 was elicited; for story recall, an estimate of the number of remembered propositions or ideas out of 100 was elicited (the number 100 having been chosen instead of the actual number of propositions per story in order to attain percent estimates, standard across different-length stories); for fact recall, an estimate of the number of facts remembered out of 40 was elicited. For complete task wording, see Appendix D.

Targeted Comparisons. Participants were also asked to assess their performance relative to that of hypothetical selves and others to determine some specific guidelines as to the implicit theories held with regards to memory aging. Performance comparisons were obtained for word recall, story recall and fact recall for each participant with regard to each of the following: an average person their age, an average 20-year-old, themselves at age 20, and the best they could possibly do. A five-point Likert scale (ranging from 0 = *much better* to 4 = *much worse*) was used. For complete task wording, see Appendix E.

## Results

### Overview

The data analyses consist of five main parts. The first part examines the profile of memory beliefs in the sample. Memory beliefs, as expressed through retrospective evaluations, numerical postdictions and targeted comparisons, are depicted. The second part examines the accuracy of the numerical postdictions. The initial part of this analysis looks at whether the numerical postdictions were more or less accurate for different occasions using postdiction-performance correlations. The next part of this analysis examines whether numerical postdiction accuracy differed across age, gender and occasions using postdiction minus performance difference scores. The last part of this analysis compares the trends of postdicted with actual scores on the 3 tasks. The third part of the analyses looks at the association of memory belief variables with MSE. The initial part of this investigation of MSE relates the various memory belief variables to MSE, while controlling for the influence of background variables including self-rated health, age, years of education, gender and marital status. The next part of this investigation looks at age, occasion and MSE differences in numerical postdictions, and the interactions of these variables. The fourth part of the analyses investigates participants' use of their beliefs concerning current performance, MSE and change in memory, in generating postdicted numerical performance values. The fifth and final part of the analyses examines the differences between the selected sample of participants, who gave complete or nearly complete responses, and those participants whose response level was too low to justify their

inclusion in the selected sample. Throughout the analyses, the main tests were evaluated using an  $\alpha$  level of .05. They are followed by post hoc analyses, evaluated at a .025  $\alpha$  level to protect against type I error rates but maintain a reasonable level of type II errors.

### Profile of Memory Beliefs

Memory beliefs were examined using repeated measures MANOVAs for three clusters of variables: retrospective evaluations, numerical postdictions and targeted comparisons.

Retrospective Evaluations. Retrospective evaluations of past compared to current performance for word recall, story recall and fact recall were examined over the three retrospected occasions. The means and standard deviations for each question are reported in Table 3. A 4-way repeated measures MANOVA was computed using Task (3 levels)  $\times$  Occasion (3 levels)  $\times$  Age Group (2 levels)  $\times$  Gender (2 levels) as the independent variables, and the retrospective evaluations for all three tasks as the dependent variables. The MANOVA revealed significant interactions of Task  $\times$  Age Group, Wilks's  $\lambda = .88$ ,  $F(2, 101) = 6.92$ ,  $p < .01$ ,  $\eta^2 = .12$ , and of Task  $\times$  Occasion, Wilks's  $\lambda = .88$ ,  $F(4, 99) = 3.36$ ,  $p < .05$ ,  $\eta^2 = .12$ , plus a significant main effect for Occasion, Wilks's  $\lambda = .92$ ,  $F(2, 101) = 4.42$ ,  $p < .05$ ,  $\eta^2 = .08$ . Within the significant effects, each set of relevant means was tested using differencing contrasts to determine where task, age group and occasion differences lay.

The main effect for Occasion was tested by contrasting the means of each of the 3 occasions with each other. The occasion effect was attributable to the difference

between evaluations of 1989 ( $\underline{M} = 1.67$ ,  $\underline{SD} = .48$ ) and 1992 performance ( $\underline{M} = 1.80$ ,  $\underline{SD} = .38$ ),  $F(1) = 8.81$ ,  $p < .01$ ,  $\eta^2 = .08$ , and 1986 ( $\underline{M} = 1.66$ ,  $\underline{SD} = .63$ ) and 1992 performance,  $F(1) = 5.51$ ,  $p < .025$ ,  $\eta^2 = .05$ . The means for 1986 and 1989 were not significantly different from each other.

The Task x Age Group interaction was tested by contrasting each of the 3 pairs of tasks for the 2 age groups. The interaction was largely attributable to significant differences between the young-old and the old-old on fact recall (Y-O  $\underline{M} = 1.80$ ,  $\underline{SD} = .38$ , O-O  $\underline{M} = 1.52$ ,  $\underline{SD} = .67$ ) and word recall (Y-O  $\underline{M} = 1.61$ ,  $\underline{SD} = .58$ , O-O  $\underline{M} = 1.75$ ,  $\underline{SD} = .72$ ),  $F(1) = 11.79$ ,  $p < .01$ ,  $\eta^2 = .10$ , and fact and story recall (Y-O  $\underline{M} = 1.78$ ,  $\underline{SD} = .51$ , O-O  $\underline{M} = 1.80$ ,  $\underline{SD} = .50$ ),  $F(1) = 8.41$ ,  $p < .01$ ,  $\eta^2 = .08$ . The interaction, as depicted in Figure 1, reveals that the age effect varies as a function of task. Both age groups believed there was little change in their story recall abilities. However, the young-old viewed their past performance on word recall as better than that in the present, whereas the old-old viewed their performance on fact recall as having been better previously.

The Task x Occasion interaction was tested by contrasting each of the 3 pairings of tasks for each of the 3 pairings of years. The contrasts revealed that the interaction was largely attributable to significant differences between word and story recall between 1989 (word  $\underline{M} = 1.59$ ,  $\underline{SD} = .71$ , story  $\underline{M} = 1.75$ ,  $\underline{SD} = .55$ ) and 1992 (word  $\underline{M} = 1.83$ ,  $\underline{SD} = .59$ , story  $\underline{M} = 1.82$ ,  $\underline{SD} = .49$ ),  $F(1) = 5.22$ ,  $p < .025$ ,  $\eta^2 = .05$ , and word and story recall between 1986 (word  $\underline{M} = 1.59$ ,  $\underline{SD} = .92$ , story  $\underline{M} = 1.79$ ,  $\underline{SD} = .73$ ) and 1992  $F(1) = 6.66$ ,  $p < .025$ ,  $\eta^2 = .06$ . Differences between fact

and story recall between 1986 (fact  $\underline{M}$  = 1.58,  $\underline{SD}$  = .74, story  $\underline{M}$  = 1.79,  $\underline{SD}$  = .73) and 1989 (fact  $\underline{M}$  = 1.67,  $\underline{SD}$  = .64, story  $\underline{M}$  = 1.75,  $\underline{SD}$  = .55),  $F(1) = 4.94$ ,  $p < .05$ ,  $\eta^2 = .05$ , and between fact and story recall between 1986 and 1992 (fact  $\underline{M}$  = 1.76,  $\underline{SD}$  = .59, story  $\underline{M}$  = 1.82,  $\underline{SD}$  = .49),  $F(1) = 3.95$ ,  $p < .05$ ,  $\eta^2 = .04$ , approached significance. This interaction, as depicted in Figure 2, demonstrates that the effect of occasion varies by task. More distant past performances for word recall were viewed as having been superior to those on the more recent occasion, and opinions of fact recall tended towards this pattern, whereas story recall performances were viewed as having been much the same across past occasions.

Ancillary Analyses. It should be noted that all the retrospective performance evaluations had values of between 1 ("somewhat better 3/6/9 years ago") and 2 ("about the same" as performance in the present). In order to address the question of whether the evaluations of previous occasion performance were significantly different from evaluations of current performance, one-sample t-tests were calculated comparing the means of the evaluations to a baseline value of 2. To protect against type I error, an  $\alpha$  level of .025 was used to evaluate the tests. As shown in Table 3, all of the values significantly differed from 2, indicating that all past performances were viewed as superior to those in the present. This paints a picture of a general perception of memory decline across three previous occasions, although the magnitude of perceived change is only moderate.

Numerical Postdictions. Numerical postdictions were collected in 1995 for each of the four testing occasions across all three recall tasks. Means and standard

deviations for the numerical postdictions are reported in Table 4. These values are presented in both absolute and percent scales. The percent values were generated in order to allow comparisons between the differently-scaled tasks. A 4-way repeated measures MANOVA was computed using Task (3 levels) x Occasion (4 levels) x Age Group (2 levels) x Gender (2 levels) as independent variables and the numerical postdictions as the dependent variables. The MANOVA revealed significant effects on the overall tests for Occasion, Wilks's  $\lambda = .71$ ,  $F(3,100) = 13.62$ ,  $p < .001$ ,  $\eta^2 = .29$ , and Task, Wilks's  $\lambda = .89$ ,  $F(2,101) = 6.58$ ,  $p < .01$ ,  $\eta^2 = .12$ , and a significant interaction of Task x Gender Wilks's  $\lambda = .89$ ,  $F(2,101) = 6.34$ ,  $p < .01$ ,  $\eta^2 = .11$ .

Polynomial contrasts indicated a significant linear trend for Occasion  $F(1) = 26.75$ ,  $p < .001$ ,  $\eta^2 = .21$ . The more distant the occasion, the higher the postdiction. The Task effect was examined using differencing contrasts to test the means of the tasks overall. These contrasts revealed that story recall postdictions ( $M = 59.72\%$ ,  $SD = 23.19$ ) were significantly lower than those for word recall ( $M = 66.37\%$ ,  $SD = 13.47$ )  $F(1) = 8.54$ ,  $p < .01$ ,  $\eta^2 = .08$ , and those for fact recall ( $M = 68.03\%$ ,  $SD = 17.02$ ),  $F(1) = 13.27$ ,  $p < .001$ ,  $\eta^2 = .12$ . Word and fact recall postdictions did not significantly differ from each other.

The Task x Gender interaction was then tested by contrasting each of the 3 pairings of tasks for both men and women. It was revealed that story recall postdictions differed from those for fact recall among men and women ( $F(1) = 10.66$ ,  $p < .01$ ,  $\eta^2 = .10$ ) with men postdicting lower story scores ( $M = 54.79\%$ ,  $SD = 20.13$ ) than women ( $M = 63.79\%$ ,  $SD = 24.88$ ), and women postdicting lower fact scores ( $M$

= 64.84%, SD = 17.85) than men (fact M = 71.89%, SD = 15.26) (see Figure 3).

Overall, results indicated a belief in the decline of performance over time. In addition, story recall performance was seen as inferior to that for word and fact recall. Furthermore, while men postdicted higher fact recall scores than did women, women postdicted higher story recall scores than did men.

Targeted Comparisons. The targeted comparison questionnaire asked participants to rate their own current memory performance as compared to that of others or themselves in hypothetical situations. The scale gave the option to rate the self as much better than, better than, about the same as, worse, or much worse than the target of comparison. An examination of the means, as reported in Table 5, reveals some patterns that were consistent across all tasks. The rank order, from most favourable comparison to least favourable comparison, was the same for all tasks. Relative to the four comparison targets, the overall sample viewed themselves as performing from best to worst, as follows: (a) average peer, (b) the best I could possibly do, (c) myself at age 20, and (d) the average 20-year old. The data were analyzed in two phases: (a) the main analysis, comparing all factors in an analysis of variance model, and (b) ancillary analyses, comparing each target separately to a baseline value.

Comparative targets were examined across all tasks in a 4-way repeated measures MANOVA using Task (3 levels) x Target (4 levels) x Age Group (2 levels) x Gender (2 levels) as independent variables, and the four target comparisons for each of the three tasks as the dependent variables. The MANOVA revealed significant 2-way

interactions for Task x Gender, Wilks's  $\lambda = .94$ ,  $F(2, 101) = 3.30$ ,  $p < .05$ ,  $\eta^2 = .06$ , and for Task x Target, Wilks's  $\lambda = .79$ ,  $F(6, 97) = 4.38$ ,  $p < .01$ ,  $\eta^2 = .21$ , and significant main effects for Task, Wilks's  $\lambda = .90$ ,  $F(2, 101) = 5.44$ ,  $p < .01$ ,  $\eta^2 = .10$ , and Target, Wilks's  $\lambda = .55$ ,  $F(3, 100) = 26.88$ ,  $p < .001$ ,  $\eta^2 = .45$ . No other main effects or interactions reached significance.

Differencing contrasts were then performed to determine where main and interaction effects lay. When interpreting these results, it may help the reader to bear in mind that the scale employed for the targeted comparisons was as follows: 0 = I did much worse than the target, 1 = I did worse than the target, 2 = I did about the same as the target, 3 = I did better than the target, 4 = I did much better than the target. The main effect of Task was tested by contrasting the means of each of the 3 tasks with each other across all targets. The Task effect was largely attributable to differences between word ( $M = 1.67$ ,  $SD = .60$ ) and fact recall evaluations ( $M = 1.80$ ,  $SD = .58$ ,  $F(1) = 7.25$   $p < .01$ ,  $\eta^2 = .07$ ) and to differences between word and story recall evaluations ( $M = 1.84$ ,  $SD = .54$ ,  $F(1) = 8.73$   $p < .01$ ,  $\eta^2 = .08$ ). Fact and story recall evaluations did not differ significantly. This main effect reveals that participants generally viewed their comparative performances on word recall as inferior to those on fact and story recall.

The main effect of Target was tested by contrasting the means of each of the 4 targets with each other across all tasks. The Target effect was attributable to the difference between each comparison with an average peer ( $M = 2.11$ ,  $SD = .47$ ). Each of the other targets was considered significantly superior to the average peer: the best I

could possibly do ( $\underline{M} = 1.75$ ,  $\underline{SD} = .74$ ,  $F(1) = 15.50$ ,  $p < .001$ ,  $\eta^2 = .13$ ), myself at age 20 ( $\underline{M} = 1.65$ ,  $\underline{SD} = .78$ ,  $F(1) = 34.59$ ,  $p < .001$ ,  $\eta^2 = .25$ ), and the average 20-year old ( $\underline{M} = 1.57$ ,  $\underline{SD} = .76$ ,  $F(1) = 74.03$ ,  $p < .001$ ,  $\eta^2 = .42$ ). The Target main effect reveals that participants viewed the performances of all targets as better than the performance of an average peer.

The Task x Gender interaction was tested by contrasting each of the 3 pairs of tasks for each gender. The interaction, as depicted in Figure 4, was attributable to the difference between evaluations of word recall (men  $\underline{M} = 1.61$ ,  $\underline{SD} = .54$ , women  $\underline{M} = 1.72$ ,  $\underline{SD} = .65$ ) and fact recall (men  $\underline{M} = 1.93$ ,  $\underline{SD} = .55$ , women  $\underline{M} = 1.70$ ,  $\underline{SD} = .58$ ),  $F(1) = 6.43$ ,  $p < .025$ ,  $\eta^2 = .06$ . Evaluations of story recall (men  $\underline{M} = 1.82$ ,  $\underline{SD} = .55$ , women  $\underline{M} = 1.85$ ,  $\underline{SD} = .54$ ) did not play a significant part in the interaction. The Task x Gender interaction reveals that the task effect varies as a function of gender. Whereas men believed their comparative performance for fact recall was higher than that for word recall, women believed their performance was comparatively higher for word recall than for fact recall.

The Task x Target interaction reveals that the effect due to the comparative target varies as a function of the tasks being contrasted. This interaction was tested by contrasting each of the 3 pairs of tasks for each of the 6 pairs of targets. The contrast results are presented in Table 5. The interaction, as depicted in Figure 5, was attributable to the differences between comparisons to the average peer and those to all other targets, within fact recall evaluations contrasted with both word and story recall evaluations. No other target comparisons within any of the task contrasts were

significant. Likewise, contrasting word with story recall evaluations resulted in no significant target comparisons.

Overall, the MANOVA presents an interesting pattern of results. Whereas the main effect of Task revealed a belief in a lower-ranked performance on word recall when compared with story and fact recall, this result held only if averaged over both genders and all targets. The Task x Gender interaction revealed that this pattern was modified through the women's tendency to rank their comparative performance on word recall as slightly higher than that on fact recall, rather than professing the reverse pattern which was the case for men and for the overall main Task effect. The main effect of Task likewise did not hold when looking at the individual targets contrasted with each other. The two main effects of Task and Target were modified by each other to produce a more complicated picture. Indeed, when considering target comparisons as well as those for task, word recall performance was ranked higher than that for fact recall when comparing the self with an average peer. Furthermore, comparative views of fact and story recall performance were not considered different when averaged over all targets, but when targets were considered individually, fact recall performance was ranked lower than that for story recall when comparing the self with both an average peer and an average 20 year-old. Conversely, it was ranked higher than story recall performance when comparing the self with the best one could possibly do, and the self at age 20. The main effect pattern of word recall performance being ranked lower than that for story recall was the only one that held across all target comparisons.

Ancillary Analyses. The response category "about the same" presents itself as a baseline value with which to compare responses regarding the different targets of comparison. T-tests (reported in Table 5) were run on the value of each mean compared to this test value. They were evaluated at an  $\alpha$  level of .025. Regarding the comparison to the average peer, participants believed their performance would be slightly better than or the same as this target. In fact, only the comparison with the average peer for story recall resulted in a rating that was significantly better for the self than for the comparative target. The second-ranked comparative target was "the best I could possibly do." For word recall and story recall, the present performance was seen as being significantly worse than the participant's own best possible performance, whereas for fact recall, the present performance was seen as being equivalent to the participant's own best performance. It appears that participants believed that with regard to the recall of facts, they were already as skilled as they would ever become, even with practice or training. The third- and fourth-ranked comparative targets were "myself at age 20" and the "average 20-year old." For all tasks, present performance was seen as significantly worse than the hypothetical performance of one's self at age 20, or the average 20-year old. Paired means t-tests within each task revealed no significant differences between "myself at age 20" and the "average 20-year old," indicating that the two hypothetical targets' performances were viewed as equivalent in comparison to participants' present performance.

#### The Accuracy of Numerical Postdictions

Occasion Differences in Accuracy as Assessed by Correlations. Differences in

the accuracy of numerical postdictions across retrospective occasions were assessed by correlating them with the actual performances for word recall, story recall and fact recall. These Pearson Product Moment correlations are presented in Table 7. Overall, the magnitude of the correlations is substantial, ranging from  $r = .41$  to  $r = .87$ , with some variation across tasks and occasions. The correlations were examined to determine whether accuracy varied systematically by occasion. Specifically, the hypothesis that more recent postdictions would be significantly more accurate than earlier ones was tested by comparing the past accuracy correlations to previous correlations within each task (see Table 7). The procedure adopted for the comparison of these accuracy values was as follows: The two correlation coefficients were transformed using the Fisher Z-transform formula ( $Z_f = 1/2 * \ln((1+R)/(1-R))$ ). The difference  $z = (Z_{f1} - Z_{f2}) / \text{SQRT}(1/(N_1 - 3) + 1/(N_2 - 3))$  is distributed approximately Standard Normally. The level of significance was determined using the z value. It should be noted that accuracy for 1995 postdictions was calculated two ways: a) postdiction correlated with the average of the scores for both versions of each task, and b) postdiction correlated only with the version of the task immediately preceding the postdiction. Method (a) was the technique used to calculate accuracy correlations for all retrospective occasions, as the questionnaire wording asked for a numerical postdiction for the task generally (rather than version-specifically) 3, 6 and 9 years ago. Method (b) was additionally used for 1995 in particular because the wording of the numerical postdiction questionnaire implied that an estimate was sought for the test version immediately preceding it, rather than the average of the two test versions.

Method (a) was employed in addition to method (b) in order that strictly parallel comparisons could be made between concurrent and retrospective accuracy. For all tasks, no significant differences emerged between measures of 1995 accuracy calculated using (a) two tests or (b) one, therefore, the practice of reporting only the two-test averages is adopted here.

The correlations between 1995 postdicted and actual performance were compared to the 3-, 6- and 9-year (retrospective) postdiction-performance correlations within each task. An examination of the significant values resulting from the comparison of 1995 accuracy with that for each previous occasion reveals the following: 1) for word and fact recall, all retrospective postdictions were significantly less accurate than concurrent ones, and 2) for story recall, all retrospective postdictions were equally accurate.

In addition to the comparison of concurrent postdictions with retrospective ones, retrospective postdictions were compared to each other for accuracy. For word recall, the accuracy correlations for 1986 and 1989 can be considered equivalent. Accuracy for 1992 was significantly greater than that for 1989 ( $p < .01$ ). In addition, accuracy for 1995 was significantly higher than accuracy for 1992 ( $p < .05$ ). Thus, postdiction accuracy within the word recall task was greater for the more recent than the earlier occasions.

For fact recall, accuracy levels for 1986, 1989 and 1992 can be considered equivalent. However, as previously noted, accuracy for 1995 was significantly higher than accuracy for the 3 earlier occasions. Thus, accuracy within the fact recall task

was greater for the concurrent (1995) postdiction than for earlier ones.

The story recall task displayed a different pattern from either the word or fact recall tasks. Postdiction accuracy did not significantly differ among any of the occasions. All correlations ranged between  $r = .41$  and  $r = .50$ .

In order to assess the direction of inaccuracy evident in the correlations, the means of the difference scores between postdicted and actual performance were calculated. For ease of comparison between tasks, the values were converted to percent scales. They are presented in Table 8. An examination of the means reveals that they are consistently positive, indicating a trend across all tasks to overestimate performance.

Evaluating Accuracy by Difference Scores. Differences in the accuracy of numerical postdictions across tasks, retrospective occasions and age groups were examined in a 4-way repeated measures MANOVA, which used Task (3 levels) x Occasion (4 levels) x Age Group (2 levels) x Gender (2 levels) as the independent variables, and the percent-converted postdiction-minus-performance difference scores as the dependent variables. As was the case in the previous analysis using correlations, the MANOVA used the mean of the two 1995 performances within each task to calculate the 1995 difference score. The MANOVA revealed significant main effects for Task, Wilks's  $\lambda = .43$ ,  $F(2, 101) = 66.92$ ,  $p < .001$ ,  $\eta^2 = .57$ , and Occasion, Wilks's  $\lambda = .83$ ,  $F(3, 100) = 7.06$ ,  $p < .001$ ,  $\eta^2 = .18$ .

The percent overestimation values for the 3 tasks were then contrasted with each other to determine where the main effects of Task lay. Task effects were due to

the differences between all 3 of the tasks with each other: word/story recall  $F(1) = 84.16$ ,  $p < .001$ ,  $\eta^2 = .45$ , story/fact recall  $F(1) = 13.08$ ,  $p < .001$ ,  $\eta^2 = .11$ , and word/fact recall  $F(1) = 70.03$ ,  $p < .001$ ,  $\eta^2 = .41$ . The mean percent overestimation values reported in Table 8 reveal that performance estimation was least accurate for story recall (21.37% to 26.05% overestimated), followed by fact recall (14.40% to 15.98% overestimated) and word recall (2.20% to 8.03% overestimated).

In order to assess the Occasion effect, polynomial contrasts of the 4 occasions were conducted upon the percent overestimation values across the 3 tasks. Figure 6 depicts the occasion effect of accuracy. The linear trend of decline in overestimation ( $F(1) = 9.78$ ,  $p < .01$ ,  $\eta^2 = .09$ ) that is evident toward the more recent occasions is moderated by a cubic trend ( $F(1) = 9.32$ ,  $p < .01$ ,  $\eta^2 = .08$ ), characterized by the sharp drop evident between 1989 and 1992.

Taken together, the results of the analyses using both correlations and difference scores reveal that concurrent postdictions tended to be more accurate than those for previous occasions. Each task revealed its own pattern of accuracy, with story recall not varying across occasions, and word and fact recall revealing greater overestimation for past occasions. In addition, performances were overestimated at different levels depending on the task in question. Story recall estimation was the least accurate, followed by fact recall, and word recall.

Comparing Postdicted with Actual Scores. In order to completely depict participants' accuracy in estimating past performance, Figures 7, 8 and 9 illustrate both the postdicted and the actual scores for word recall, story recall and fact recall

respectively. It is of interest to observe the patterns found across the two types of scores. A glance at each of the postdicted trajectories reveals the participants' belief in a trend of linear decline noted above in the numerical postdictions profile section. A 1-way repeated measures MANOVA was conducted on the 4 levels of Occasion, using the numerical postdictions for each separate task as the dependent variables. As with the previously conducted postdiction MANOVA (see above section on the profile of memory beliefs), the present MANOVA revealed a significant effect for Occasion, Wilks's  $\lambda = .60$ ,  $F(9,97) = 7.29$ ,  $p < .001$ ,  $\eta^2 = .40$ . Unlike the previously conducted MANOVA, the present analysis has the ability to highlight differences between tasks in the patterns of postdictions over occasions. Toward this end, polynomial contrasts were conducted on the means of each of the postdicted occasions within all 3 tasks. Results revealed a significant linear trend within each task (word:  $F(1) = 7.78$ ,  $p < .01$ ,  $\eta^2 = .07$ ; story:  $F(1) = 15.01$ ,  $p < .001$ ,  $\eta^2 = .13$ ; fact:  $F(1) = 34.02$ ,  $p < .001$ ,  $\eta^2 = .25$ ), as well as a significant cubic trend within word recall,  $F(1) = 6.66$ ,  $p < .025$ ,  $\eta^2 = .06$ . These patterns for word, story and fact recall are depicted in Figures 7, 8 and 9 respectively.

Also illustrated in these figures are the means for the actual performances for each of these tasks. An examination of these actual performance trajectories indicates a slightly more complicated picture than did the postdictions, as slight variations in performance are evident for word recall and story recall, and a straighter slope of decline is revealed for fact recall. To assess the performance trajectories, a 1-way repeated measures MANOVA was conducted on the 4 levels of Occasion, using the

performance levels for each separate task as the dependent variables. The overall effect of Occasion was significant, Wilks's  $\lambda = .63$ ,  $F(9,97) = 6.23$ ,  $p < .001$ ,  $\eta^2 = .37$ . The 3 tasks were then examined using polynomial contrasts on the performance means for each occasion within the task. Only one of the tasks, fact recall, revealed a significant linear trend of decline,  $F(1) = 31.79$ ,  $p < .001$ ,  $\eta^2 = .23$ , thus demonstrating that the participants' beliefs about the trend for this task were accurate. The other two tasks demonstrated different performance patterns. Story recall revealed a significant cubic trend,  $F(1) = 5.16$ ,  $p < .025$ ,  $\eta^2 = .05$ , and word recall demonstrated no significant trends. A glance at Figure 8 reveals that while performance for story recall went up and down over the occasions of testing, it remained much the same from the initial testing occasion to the current one. Figure 7 reveals that performance for word recall also went up and down a little, but overall did not conform to the linear pattern of decline anticipated by the sample.

The combined results of the postdiction and performance MANOVAs present an interesting pattern. While the patterns of postdictions reveal the sample's belief in their linear decline over time in all the tasks, the patterns of performance reveal patterns of both decline and maintenance over time, depending on the task in question.

#### The Association of Memory Beliefs with MSE

Partial Correlations of Memory Belief Variables with MSE. In order to assess whether MSE was associated with the three sets of memory belief variables (retrospective evaluations, targeted comparisons and numerical postdictions), the correlation of each memory beliefs variable with MSE was calculated for all tasks and

all occasions. These correlations were computed partialling out self-rated health, age, years of education, gender and marital status, in order to control for their influence. It is considered that each of these variables may share variance with MSE and that controlling for them will generate a clearer pure picture of MSE's unique influence. Significant results included the following retrospective evaluations: story recall for 1989 compared to 1995 ( $\underline{pr} = .24, p < .05$ ), and for 1992 compared to 1995, ( $\underline{pr} = .20, p < .05$ ), and the following targeted comparisons: for fact recall, myself compared to an average peer, ( $\underline{pr} = .21, p < .05$ ), for story recall and word recall, myself compared to an average 20-year-old, ( $\underline{pr} = .28, p < .01$ , and  $\underline{pr} = .24, p < .05$  respectively), and for word recall, myself now compared to myself at age 20, ( $\underline{pr} = .26, p < .01$ ).

Very few of the memory belief variables were associated with MSE, and the ones which were had low correlations ( $\underline{pr} = .20$  to  $.28$ ). Interesting to note are the absence of any significant correlations between MSE and the numerical postdictions. The retrospective evaluations variables which were correlated with MSE were the comparisons of story recall for the two most recent occasions. Lower MSE was associated with a worse conception of current performance compared with that in the past. MSE could be partially influencing the assessment of the relative strength of current compared to past performance for this particular task. The targeted comparisons also produced no consistent patterns. However, the variables associated with lower MSE reflected a conception of the self as performing poorly compared with the target. When comparing themselves to an average 20-year-old for both word and story recall, participants may have been influenced by their current MSE.

Age, Occasion and MSE Differences in Postdictions. Hierarchical multiple regressions were performed, entering Age, Occasion and MIA Capacity (the representation of current MSE) in the first block, and the interactions of Age and Occasion, Age and MIA Capacity, and MIA Capacity and Occasion, in the second block, and the three-way interaction of Age, Occasion and MIA Capacity in the third block. The dependent variable was either word recall postdiction, story recall postdiction or fact recall postdiction. The variables which predicted the postdictions, and the details of the overall model for the block of variables in which the variables were significant, are recorded in Table 9. For each of the 3 tasks, a slightly different pattern emerged. For word recall, age and occasion predicted postdictions. For story recall, age, MIA Capacity and the interaction of the two predicted postdictions. For fact recall, occasion and MIA Capacity predicted postdictions.

Higher word postdictions were predicted by lower ages and earlier occasions. Higher story postdictions were predicted by higher ages, higher MSE, and lower age x MSE interactions (the lower the age, the higher the MSE). Note that the unique variance of both age and MSE in this case was low. The  $R^2$  for the first block of variables was only .010. The  $R^2$  value increased by .052 when the interactions were entered in the second block. Thus, the prediction of story postdictions can be largely attributed to the interaction of age and MSE. It is notable that age and MSE are moderately correlated ( $r = -.30$ ,  $p < .01$ ). The other notable feature of the model was that higher fact recall postdictions were predicted by earlier occasions and higher MSE.

### Predicting Past Postdiction using Current Postdiction and Change

To investigate whether ideas of current performance and implicit beliefs about memory change were systematically employed in order to generate retrospective performance evaluations, the prediction of past numerical postdictions was assessed using the current postdiction, MSE and MIA Change. Separate hierarchical multiple regressions were performed for each task, entering the 1995 numerical postdiction (considered to be the participant's representation of current performance) at the first step, MIA Capacity (a representation of current MSE) at the second step, and MIA Change score (a representation of implicit beliefs about the recent trajectory of memory change) at the third step, using past numerical postdiction (either 1986, 1989 or 1992) as the dependent variable. Results are presented in Table 10.

To summarize, results for all tasks and all occasions indicated that the current postdiction consistently predicted the past postdiction. When added to the model on the second step, MIA Capacity did not predict past postdiction for any of the tasks. This indicates that current MSE was not combining with current postdiction to predict the assessment of past performance. Interestingly, when MIA Change was added to the model on the third step, MIA Capacity predicted the word recall postdiction for 1986. This unique finding tells us that when added to current performance perceptions, MSE combined with memory change beliefs to predict the most distant retrospective performance postdiction for word recall. A more general pattern of results was also generated when MIA Change was added to the model on the third step. It significantly predicted the 6- and 9-year postdictions for both story recall and

word recall, but none of the numerical postdictions for fact recall. This indicates that beliefs about memory change in combination with ideas about current performance predicted assessments of past performance for word and story recall, but not for fact recall.

#### Differences Between Participants in Selected and Unselected Samples

In order to determine whether a self-selection process occurred by the failure of some participants to answer multiple questions, analyses were conducted to compare the unselected participants with the selected sample used for analyses. A one-way MANOVA was computed for Sample (unselected vs. selected) using age, years of education, marital status, MSE, self-rated health and gender as dependent variables. The MANOVA revealed a significant effect for Sample, Wilks's  $\lambda = .88$ ,  $F(6, 164) = 3.85$ ,  $p < .001$ ,  $\eta^2 = .12$ . Univariate tests indicated significant sample differences on age ( $F(1) = 9.11$ ,  $p < .01$ ,  $\eta^2 = .05$ ), years of education ( $F(1) = 6.74$ ,  $p < .05$ ,  $\eta^2 = .04$ ) and marital status ( $F(1) = 7.34$ ,  $p < .01$ ,  $\eta^2 = .04$ ).

Table 11 lists the means and standard deviations for background, performance and metamemory variables for the two groups of participants. To supplement the variables pertaining to subject background, Table 12 presents the proportion of each group falling into meaningful categories of education level and marital status. It can be seen from the subject background information contained in Tables 11 and 12 that participants in the group not selected because of missing data (a) were older, (b) had attained lower levels of education, and (c) were less likely to be married than the selected sample.

Interesting to note are the low numbers of participants in the unselected sample for whom complete information is available on certain variables. The number of participants giving valid information for individual data points varies from 67 (100% of the unselected sample) for variables such as age, to 30 (44.8% of the unselected sample) for the 1986 fact recall postdiction. The variables for which data are most often missing in this subsample are the numerical postdictions. The trend within each task is one of an increased answer omission rate with increased distance of the postdicted occasion. In order to assess whether this trend was significant, a one-way repeated measures ANOVA was computed using Occasion (4 levels) as the independent variable, with the number of missing postdictions across all tasks as the dependent variables. The ANOVA revealed a significant main effect of Occasion, Wilks's  $\lambda = .58$ ,  $F(3, 64) = 15.22$ ,  $p < .001$ ,  $\eta^2 = .42$ . Differencing contrasts were conducted on the means of each pair of successive years to assess where the occasion effects lay. Results showed significant differences between the number of missing postdictions in 1989 ( $M = 1.28$ ) and 1992 ( $M = 1.10$ ),  $F(1) = 8.77$ ,  $p < .01$ ,  $\eta^2 = .12$ , and between 1992 and 1995 ( $M = .30$ )  $F(1) = 45.69$ ,  $p < .001$ ,  $\eta^2 = .41$ . The 1986 ( $M = 1.36$ ) contrast with 1989 approached significance,  $F(1) = 3.72$ ,  $p \leq .058$ ,  $\eta^2 = .05$ , and was in keeping with the pattern of increased omission with distance of postdiction.

The trend across all 3 tasks was for the unselected sample to perform more poorly than the selected sample. In order to examine whether performance levels for each recall task differed between the 2 samples, a 3-way repeated measures MANOVA, with Task (3 levels)  $\times$  Occasion (4 levels)  $\times$  Sample (2 levels) as the

independent variables, and the percent-converted performance scores for each task as the dependent variables was calculated. Sufficient data were missing for the unselected sample that the sample size used in the MANOVA was  $N = 59$ , rather than  $N = 67$ . The MANOVA revealed significant interactions of Sample  $\times$  Occasion, Wilks's  $\lambda = .93$ ,  $F(3,161) = 4.33$ ,  $p < .01$ ,  $\eta^2 = .08$ , and of Task  $\times$  Occasion, Wilks's  $\lambda = .74$ ,  $F(6,158) = 9.27$ ,  $p < .001$ ,  $\eta^2 = .26$ , plus significant main effects for both Task, Wilks's  $\lambda = .15$ ,  $F(2,162) = 450.59$ ,  $p < .001$ ,  $\eta^2 = .85$ , and Occasion, Wilks's  $\lambda = .78$ ,  $F(3,161) = 15.07$ ,  $p < .001$ ,  $\eta^2 = .22$ . The between subjects effect for Sample was also significant,  $F(1) = 6.00$ ,  $p < .05$ ,  $\eta^2 = .04$ .

The Sample effect indicates that on average, the unselected sample performed more poorly than did the selected sample. However, this effect is modified by an interaction with Occasion. As depicted in Figure 10, it is interesting to note that the unselected sample's performance in 1995 was lower than that for each of the previous occasions, while the selected sample's average performance went up and down. In addition, the unselected sample's highest average performance across tasks, achieved in 1986 ( $M = 48.16$ ), was similar to the selected sample's lowest average performance across tasks, achieved in 1995 ( $M = 48.96$ ). To assess the differing patterns of performance between the 2 samples across the different occasions, differencing contrasts were performed on the means of the combined tasks for each pair of years. As with previous contrasts, they were evaluated at an  $\alpha$  level of .025. The gap between the performances of the 2 samples were found to be greater for 1995 than for 1986 ( $F(1) = 45.47$ ,  $p < .001$ ,  $\eta^2 = .22$ ), and 1989 ( $F(1) = 12.83$ ,  $p < .001$ ,  $\eta^2 = .07$ ).

Furthermore, the difference between 1995 and 1992 approached significance ( $F(1) = 5.08, p < .05, \eta^2 = .03$ ), as did that between 1992 and 1989 ( $F(1) = 4.30, p < .05, \eta^2 = .03$ ). To summarize this trend, on each of the past occasions, the gap between the performances of the selected sample and the unselected sample was smaller than that on the current occasion. The most dramatic difference between the 2 samples was found in 1995.

The 2 samples' numerical postdictions values over retrospective and concurrent occasions are also reported in Table 11. With the exception of word recall in 1986, the trend across all tasks and occasions was for the unselected sample to postdict lower scores than the selected sample. In order to assess whether the 2 samples differed significantly in the values postdicted, a 3-way Task (3 levels) x Occasion (4 levels) x Sample (2 levels) repeated measures MANOVA was conducted, using the means of the 3 tasks for each occasion as the dependent variables. Sufficient data were missing for the unselected sample that the sample size used in the MANOVA was  $N = 15$ , rather than  $N = 67$ . The small sample size resulted in insufficient power to detect a group effect, so the analysis was re-done without the Task factor, using the averaged percent-converted means of all three tasks as the dependent variables. This resulted in a sample size of  $N = 49$  out of the possible 67, a much more powerful test of the Sample effect. The 2-way Occasion (4 levels) x Sample (2 levels) repeated measures ANOVA revealed a significant main effect for Occasion, Wilks's  $\lambda = .77, F(3,151) = 14.82, p < .001, \eta^2 = .23$ . The test for a between subjects effect found a significant difference for Sample,  $F(1) = 4.92, p < .05, \eta^2 = .03$ . This effect is

depicted in Figure 11. The tendency of the unselected sample to postdict lower scores may be indicative of either a belief in or an awareness of lower performance levels. Likewise, the selected sample's trend to postdict higher scores could indicate either its belief in or awareness of higher performance levels. Different accuracy in postdictions across the two samples would indicate lesser awareness on the part of one of the groups, and, for that group, would underscore the importance of memory beliefs over awareness in the production of numerical postdictions for past performances. Equivalent accuracy in postdictions would signal that performance awareness played an equal role in the generation of postdictions for both groups.

Accuracy differences between the selected and the unselected samples were examined using a 3-way repeated measures MANOVA computed for Task (3 levels) x Occasion (4 levels) x Sample (2 levels), with the postdiction minus performance difference scores for each task on each occasion as the dependent variables. Sufficient data were missing for the unselected sample that the sample size which resulted in the MANOVA was  $N = 11$ , rather than  $N = 67$ . This created a need to adopt the same strategy as the postdiction ANOVA, with the Task factor being dropped, and the 3 tasks being combined into average percent overestimation scores for each occasion. This resulted in an unselected sample size of  $N = 48$ . A 2-way Occasion (4 levels) x Sample (2 levels) repeated measures ANOVA was thus computed. The test for Sample failed to reach significance ( $F(1) = 3.19$ ,  $p = .076$ ,  $\eta^2 = .02$ ), indicating that accuracy was equivalent between the two groups.

Thus, it appears that the selected sample performed better than the unselected

sample and the groups were equally aware of their respective performance levels. Since accuracy was not different between the groups, it is likely that they each used the same strategy in generating their postdictions. For example, if either group weighted more heavily their age-decline memory beliefs in producing postdictions for past occasions an inflated inaccuracy level would have resulted for that group. This consistency between the samples therefore argues in favour of the use of the same general strategy for generating postdictions by both groups.

## Discussion

In general, this study was intended to increase the overall understanding of memory beliefs in the elderly. More specifically, it assessed profiles and accuracy of memory beliefs, as well as age, MSE, and other potential influences. The sample consisted of community dwelling adults living in a medium-sized metropolitan area. They ranged in age from 63 to 92 on the most recent occasion, and were measured at 3-year intervals over a 9-year period, yielding 4 occasions of measurement.

### Profile of Memory Beliefs

The first phase of analyses focused on the profile of memory beliefs in the elderly. Consistent with past research which revealed a belief in memory decline in old age (e.g. Heckhausen et al., 1989), the hypothesis that the elderly would believe their memory has generally declined was supported in the present study. Each set of memory belief variables depicted an overall belief in the decline of memory skills over time.

The retrospective evaluations revealed that participants believed past performances were superior to the most recent ones. Although this endorsement of aging-related decline was applied to all three tasks, the pattern of perceived decline differed by task. Specifically, the decline functions were steeper for word and fact recall than for story recall. For the latter, all past performances were seen as being much the same, yet superior to present performance.

There was a Task x Age Group interaction which revealed that the young-old adults perceived their past word recall performance as better than that in the present

whereas the old-old did not. In contrast, the old-old adults, but not the young-old, perceived this pattern for fact recall. This interaction suggests that the young-old adults may be perceiving an initial decline in word recall, whereas the old-old adults may have already passed such a stage in their perception, and now see decline for this task as having levelled off. In addition, the old-old may now be perceiving a decline in their ability to recall facts, an ability which the young-old perceive as being better maintained.

The numerical postdictions revealed some interesting patterns. When the postdictions were translated to percent scales for ease of comparison, it was revealed that different tasks were considered to have been performed at different levels. Specifically, story recall postdictions were lower than those for fact or word recall, which were postdicted at the same level. This could mean that respondents believed they performed relatively "worse" on story recall than on fact and word recall. It is entirely plausible that story recall is considered inherently more difficult than are fact or word recall. Respondents may have believed that they could not possibly remember as many as the 100 ideas they were informed were contained in each story. However, the phenomenon could also be a function of the difficulty of estimating performance on story recall. A number of factors may figure in this. One factor is the difficulty of understanding what is meant by an idea, which was defined as follows on the numerical postdiction questionnaire: "concepts [indicated by words or phrases] that are relatively important to the substance of the story." A second factor is the difficulty of quantifying how many ideas were recalled, when the form of recall was in

sentences or phrases. The idea that story recall is considered inherently more difficult than word or fact recall is borne out by its lower postdictions, and evidence for its greater difficulty of estimation comes from their lower accuracy. These between-task accuracy differences will be further discussed later.

Interesting though the Task effect is, it is modified by Gender for the numerical postdictions. The Task x Gender interaction revealed that men postdicted lower story recall scores than did women, while women postdicted lower fact recall scores than did men. Word recall postdictions did not significantly differ between the sexes. Thus, performance on the story and fact recall tasks was viewed more or less favourably depending on gender.

The differences between assessments of comparative targets helped further depict the profile of beliefs about memory held by the sample. For all tasks, a same-aged peer was ranked as the worst-performing target, followed by the self's best possible performance, with the best performances being attributed to the self at age 20 and an average 20 year-old.

The same-aged peer was considered inferior to each of the other targets. In addition, the same-aged peer was considered similar to the self on fact and word recall, and worse than the self on story recall. Past research has found that the elderly reported more age-related problems for other members of their cohort than for themselves (e.g. Crockett & Hummert, 1987). This tendency can help explain the sample's opinion that they performed better than a same-aged peer on story recall. The depiction of the sample's contemporaries translates to a relatively low conception

of their memory compared with that of the other targets.

The low rating of the self's current performance compared to the self's best possible performance for word and story recall depicts the belief that the current performance for these tasks was lower than abilities would dictate. It was believed that the performance given by the individual on these tasks could be improved given practice or training. However, the rating of fact recall performance did not reflect this belief, thus preventing the depiction of a general schema of plasticity across all tasks.

The high estimation of the memory performance capabilities of people at age 20 compared to the self at present for all tasks depicts a belief in the decline of capabilities for the tasks with age. This pattern is similar to that depicted by McFarland et al. (1992), who found that the elderly believed their memory was superior when they were younger.

While all three clusters of memory belief variables triangulated in their depiction of a general belief in decline with age, there were some differences between the clusters of variables. Specifically, the pattern of interactions within the retrospective evaluations and the postdictions differed, with the retrospective evaluations revealing interactions for Task x Age Group and Task x Occasion, and the numerical postdictions revealing an interaction for Task x Gender. These types of variations are what one might expect between multiple indicators of a complex phenomenon like implicit theories of memory. For example, the way in which thinking was structured differed between the retrospective evaluations (which called for a comparison of past with present performance) and the postdictions (which called

for absolute evaluations for all occasions). The practice of thinking comparatively versus absolutely may have prompted different responses among the respondents. Thus, rather than indicating inconsistencies in participants' ideas, the different interactions resulting from the separate clusters of memory belief variables can be seen as having tapped different aspects of the beliefs, thus providing a more rounded picture of the implicit theories of memory they held.

One particular aspect of the retrospective evaluations and the postdictions which differed was the type of scale employed. The retrospective evaluations involved the use of a Likert scale which allowed only crudely graded performance estimates, whereas the postdictions used a continuous scale which permitted more finely graded assessments. In addition, the response categories of the Likert scale are likely to have held different meanings for different individuals and possibly for the different age groups and genders among the respondents. For example, a decline in word recall performance of 3 words out of 30 over the 9-year period may mean to one participant he/she performed "much better" previously, to another that he/she performed "better" in the past, and to a third that he/she did "about the same." If the perceived change necessary for most respondents to endorse the responses depicting difference from the present was quite pronounced (e.g., 6 words) then these respondents are only likely to have endorsed a differences response if such a severe change was perceived. However, the same respondents may have depicted belief in a difference from past to present by postdicting slightly higher scores for past occasions. Thus, the more specific nature of the continuous scale employed for the numerical postdictions may

have elicited different responses from individuals and groups than did the retrospective evaluations.

Overall, the sample revealed a system of beliefs in age-related memory decline. In the terminology associated with the implicit theory hypothesis, this belief in decline would be considered evidence that a theory of difference from the past was operating with regards to memory performance. Whether this belief in decline was accurate is another question.

#### Accuracy of Memory Assessments

Much memory assessment accuracy research in the past was cross-sectional and focused on such issues as age differences in the accuracy of performance predictions (e.g. McDonald-Miszczak, Hunter & Hultsch, 1994) and the accuracy upgrade of postdictions over predictions (e.g. Hertzog et al., 1994). The present study was interested in whether participants were accurate in their specific views of the trajectory of their performance over time. The hypothesis that numerical postdictions for more recent occasions would be more accurate than those for more distant ones was supported for word and fact recall, but not for story recall. The postdictions for word and fact recall were most accurate for the concurrent occasion, but significantly less so for the retrospective ones, whereas the postdictions for story recall were equally accurate across all occasions.

Each task resulted in a unique range of accuracy scores. The peculiarity of these scores to each task may be attributable to the ease or difficulty of performance estimation for the tasks themselves. For example, the word recall task resulted in the

greatest accuracy. This is likely due to the ease of calculating how many words were remembered in the test immediately preceding the numerical postdiction, then extrapolating to the previous years' performance based on the implicit theory of consistency or difference adopted. Particularly helpful to this process may have been the sequentially numbered lines upon which participants wrote the words they recalled, which, if attended to, amounted to an approximate running performance tally during task execution. It should be noted that there is no evidence that this cue was actually attended to. In contrast, the story recall task resulted in lower accuracy. This could have been due to the difficulty of estimating performance on even the recently completed test. As noted earlier, a number of factors may figure in this, including the difficulty of understanding what is meant by an idea, and the difficulty of quantifying how many ideas were recalled. One notable point is that the lower accuracy of the story recall postdictions may be more reflective of the level which is typical for everyday retrospections, given that there are likely to be similar ambiguities operating for both. It is likely, then, that the accuracy of the story recall postdictions represents the most ecologically valid measure of everyday retrospective accuracy, and hence, the accuracy of implicit theories of memory decline.

The accuracy correlations for all the tasks for the past occasions were fairly substantial ( $r = .41$  to  $.69$ ). For 2 of the 3 tasks, the accuracy of participants' specific views of their performances was worse for past occasions than for the current one (see Table 7). This supports the idea that past performance levels were more ambiguous than those in the present. For past performance levels to have been estimated as

accurately as those in the present would have necessitated that they be monitored in some way at the time of task completion, and the past experience of monitoring recalled currently and accurately. While participants may have monitored their performance at the time of task completion and so generated an approximate feeling of how well they did, there is no evidence that they attempted to quantify their performance in the manner of an "internal postdiction." Such a performance tally might have been reasonably accurate, judging by the current study's concurrent postdiction-performance correlations, which ranged from .50 for story recall to .83 for word recall. Even had they generated concurrent performance tallies, it seems highly unlikely that they would have maintained such estimates in memory for 3 to 9 years. Without such quantifications to draw upon, participants would have had to rely on their memories of actually performing the task 3, 6 and 9 years ago, perhaps trying to remember how competent they felt on the days they came for testing, and how they felt after completing each of the tests. Based on such memories, they would then have generated quantified postdictions for the first time in 1995. Given that participants completed from 21 to 28 tasks during the 3 to 4 testing sessions of each occasion, and that the tasks were performed up to 9 years ago, for participants to have remembered either their thoughts and feelings surrounding the individual tasks in question on each occasion, or an internally postdicted estimate, seems highly unlikely. Future longitudinal research requiring postdictions at each occasion of testing would be required in order to test whether concurrent performance estimation assists in generating retrospective estimates. There is a far more plausible explanation of the

way in which participants generated performance estimates for past occasions: Ross' (1989) implicit theory hypothesis.

Given that estimates of past performance were fairly good, they were still notably inferior to those given for the present. This provides evidence that past memory performance estimates were not clearly remembered. Ross' hypothesis predicts that the conception of current performance, combined with an implicit theory of consistency or difference, would have been used to generate a picture of past performance levels. Because (a) the overall pattern of accuracy correlations is consistent with past performances not being clearly remembered, and (b) the adoption of a difference/decline principle was illustrated through the profile of memory beliefs depicted earlier, and (c) the most likely alternative explanation (depicted above) seems highly implausible, the data for word and fact recall support the idea that participants used an implicit theory of decline in order to generate their postdictions for past performances.

In addition to the patterns of change espoused by participants being of systematic decline, the direction of the perceived change was biased in a consistent manner. Difference scores revealed that the magnitude of decline was consistently exaggerated for each of the tasks. That is, participants overestimated their earlier performances more than their more recent performances. This is again consistent with the operation of an implicit theory of the decline of memory. Had the estimates been arrived at in a non-systematic (random) manner, they would have been evenly distributed about the mean, a pattern most definitely not evident in the present data.

An important part of the implicit theory hypothesis is that current status is used as a guide to generate an estimate of past status, along with a theory of consistency or difference (Ross, 1989). If a theory of consistency were operating, past estimates would be the same as current ones. However, the postdicted values for the past occasions were, without exception, higher the more distant the occasion. This pattern is entirely consistent with the employment of the current performance estimate as a benchmark, combined with a theory of difference (decline) from past occasions to produce ideas about memory performance in the past, which is the essence of the implicit theory hypothesis of beliefs about the past self.

The evidence with regard to performance levels over time suggests that the theory of difference/decline with regard to memory beliefs may be valid for some tasks but not for others. Fact recall conformed to the sample's expectation of linear decline over time, whereas story recall and word recall did not. Indeed, story recall performance did not decline over the 4 occasions, but displayed only slight variability across occasions. Word recall likewise did not display a linear decline, but varied slightly across occasions. The similarities and differences between patterns of postdictions and patterns of performance tell us that participants were more accurate in their views of the trajectory of their capabilities for certain tasks than for others.

Whereas the tasks varied in the extent to which they conformed to the participants' specific views of decline patterns, the overestimation of decline for all 3 tasks over the 4 occasions was negligible. It is interesting to note that the magnitude of percent decline depicted by the postdictions was very small. As calculated by

percent recall postdicted at Time 1 minus percent recall postdicted at Time 4: word recall decline was depicted at 3.9%, story recall decline at 4.6%, and fact recall decline at 5.9%. Given that the actual decline for this period was only 2.5% for word recall, .05% for story recall, and 4.6% for fact recall, the sample's estimated decline exceeded the actual decline by only 1.4% for word recall, 4.55% for story recall, and 1.3% for fact recall. When presented this way, the sample's estimates of decline seem reasonably accurate. That the elderly are aware of their memory decline is consistent with the findings of Johansson et al., (1997) who reported that actual decline was associated with retrospectively-reported decline on 3 of 5 memory tasks, and with McDonald-Miszczak et al. (1995), who reported that correlations were observed between longitudinal changes in memory performance and retrospective perceptions of memory change.

This evidence would suggest that the level of memory loss awareness which exists could be helping drive the process of past memory status recollection. However, the evidence is not strong enough to warrant a view of awareness and implicit theories operating to each other's exclusion. The information available so far reveals an awareness of decline which is of a general, rather than a specific and precisely accurate, nature. The two mechanisms, general awareness and implicit theories, are easily reconcilable. The general awareness of memory loss would not necessitate that the postdictions be precisely accurate because awareness at this level could serve to reinforce the memory decline belief and thus contribute indirectly, rather than directly, to the postdiction. If the elderly are generally aware that they

have experienced memory decline, it seems reasonable to expect that this awareness would contribute to the generation of postdictions in this way.

### The Association of Memory Beliefs with MSE

The hypothesis that higher MSE would be associated with the perception of memory maintenance, and lower MSE would be associated with the perception of decline was only moderately supported in the present study.

Current MSE was not highly correlated with any particular cluster of memory belief variables, but was moderately associated with some of the variables. These few associations were in the expected direction: a higher belief in the ability to use memory effectively in demanding situations was associated with a perception of memory maintenance, and a lower MSE level was associated with a perception of decline. However, so few of the variables displayed this pattern that it is not considered a strong one.

One surprising result was the absence of any significant partial correlations among MSE and the postdictions. One factor which may have contributed to this lack of relationship is the role of task experience. As Hertzog et al. (1994) point out, when lacking familiarity with a specific performance situation, individuals may utilize their global MSE to predict their performance level. An illustration of this is Hertzog, Dixon and Hultsch's (1990b) finding that the correlation of MSE with performance prediction declined slightly across trials. This could have resulted from the initial prediction being based more heavily on MSE than on first-hand task knowledge, and subsequent appraisals relying more heavily on task experience as a guideline. A

parallel result was obtained by McDonald-Miszczak et al. (1994), who used a regression analysis to relate MSE to performance predictions. They found that general MSE beliefs were used to generate initial predictions, whereas task specific knowledge was used to generate subsequent ones. To apply this principle to the present study, which employed a postdiction paradigm whereby task familiarity was already established at the time of performance estimation, the influence of global MSE would likely have been reduced prior to the generation of performance estimates, thus helping to account for the lack of a relationship. It is notable that all the clusters of memory belief variables were measured after participants gained task experience, so that the same principle may have attenuated MSE's relationship with them. Through the patterns depicted above, it is clear that although current MSE is part of the system of beliefs about one's memory capabilities, it is not the only factor playing a part.

While the correlational analysis failed to reveal a relationship between MSE and postdictions, a hierarchical regression involving Age, Occasion, and MSE, revealed significant relationships between MSE and 2 of the 3 tasks. Higher MSE was associated with higher postdictions for story and fact recall. This is the pattern that was predicted. However, the effect sizes in both cases were quite small, with Age and MSE together accounting for 5.2% of the variance in story recall, and Occasion and MSE together accounting for 3.5% of the variance in fact recall. In the case of story recall, higher MSE and lower age also interacted to produce higher postdictions. This pattern is not surprising, considering previous findings that lower age is associated with higher MSE (see Berry & West, 1993). Higher word postdictions were

associated with lower ages and earlier occasions, but surprisingly, not with MSE. The present data would suggest that postdictions for certain memory tasks are more likely related to global MSE than others. Global MSE, as measured by the MIA Capacity subscale, may be moderately associated with story and fact recall postdictions, and not with word recall postdictions. Consistent with the ideas of Berry and West, it is possible that the MIA Capacity subscale was not specific enough to any of the tasks to produce strong relationships between postdictions and MSE.

Overall, there were moderate relationships between memory belief variables and MSE. Two explanations may help account for the moderate nature of these relationships. First, task experience in each of the clusters of memory belief variables leads to the reduced influence of MSE. Second, the global nature of the MIA Capacity subscale leads to an attenuated relationship between the tasks and MSE.

#### Predicting Past Postdiction using Current Postdiction and Change

The implicit theory hypothesis for memories about the self depicts people as generating estimates of their past status for attributes based upon (a) an estimate of their current status, and (b) a theory of consistency with or difference from the past (Ross, 1989). The question investigated in the present study was whether respondents generated their past performance levels by using current memory performance beliefs combined with ideas about memory change. Hierarchical multiple regressions were performed to determine whether the current postdiction together with the Change scale of the MIA would predict past postdictions. The hypothesis that the two would combine to predict past status estimation was supported for word and story recall, but

not for fact recall.

It should be noted that the unique variance accounted for by MIA Change varied between tasks (word recall = 10% in 1986, 4% in 1989; story recall = 2% in 1986, <1% in 1989). The greater proportion of variance accounted for in the word recall postdictions suggests that memory change beliefs were more strongly related to this task than they were to story recall. Interestingly, story recall performance showed no significant variation over the 4 occasions of testing (actual decline = .05%), leading to the speculation that participants were aware that an age-decline theory of memory does not apply to this task. An equivalent awareness may not have been present for fact recall performance, or results might have revealed a significant association between fact recall and memory change, which actually declined by 4.6%. The patterns found suggest that there is a general memory decline theory, but that it may be more refined and differential than might have been expected.

Fact recall may not have followed the same pattern as word and story recall partly because the MIA Change scale is too concentrated on evaluating episodic memory change beliefs. Participants may have differentiated this from their beliefs about a semantic task. Although reliable and factorially valid (e.g. Dixon & Hultsch, 1983), the MIA Change scale was designed to provide a general measure of memory change beliefs. Inspection of its individual questions reveals that it could possibly be divided into items that tap beliefs about memory in general, as well as episodic memory and semantic memory separately. General items constitute 44% of the scale's questions, episodic items constitute at the very least 28% of the questions, whereas

semantic items may constitute at the very most 28% of the items. Future research might separate the items into general, semantic, episodic and subscales, and assess their association with past postdictions. One might expect the resulting semantic and general subscales to predict the past postdiction for fact recall, and the general and semantic subscales to behave similarly with regard to story and word recall.

In addition, it is interesting that an association between MIA Change and numerical postdiction emerged for some past occasions and not for others. It can be argued that the more recent occasions of testing are regarded by participants as being almost part of the present, and that their beliefs regarding memory change hold in particular with regards to the distant past. A lag of 3 years may not be seen as distant enough to be associated with strong memory-decline beliefs: those are reserved for more distant occasions. The period of 6 to 9 years may be just long enough to tap into conceptions of the distant past, thus accounting for the associations of word and story recall with memory change beliefs. Even so, the fact recall postdictions were not predicted by the same memory conception variable (MIA Change) that successfully predicted word and story recall postdictions.

Overall, there was a variable correspondence of the different tasks to the predicted pattern. In the case of fact recall, some potential explanations include that task-specific memory theories, an "episodically concentrated" memory change scale and a task-specific critical time lag factor are contributing to the different pattern followed. Future research is needed in order to clarify such things as whether memory theories are task-specific, whether memory change scales are more predictive when

divided into question types, and how long a period of time needs to pass in order to tap memory change beliefs.

### Sample Differences

The final phase of the study addressed whether there were systematic differences between participants in the selected and unselected samples. With regard to the performance measures, the two groups performed similarly on 6 of the 12 measures. These were word recall in 1992, 1989 and 1986, story recall in 1989 and 1986, and fact recall in 1989. The selected sample performed better on all three tasks in 1995, on story and fact recall in 1992, and on fact recall in 1986. The average magnitudes of these sample differences were as follows: Word recall (1995) = 2.25 words of a possible 30, story recall (1995 & 1992) = 4.04 propositions of a possible 100, and fact recall (1995, 1992 & 1986) = 2.32 items of a possible 40. The existence of lower performances for all tasks in 1995 is an interesting phenomenon, as it may signify a slightly accelerated memory decline trend among participants in the unselected sample compared to those in the selected sample. The reluctance of the unselected sample to offer performance evaluations may stem from an awareness of this decline. With regard to the retrospective evaluations, the two samples were equivalent on 8 of the 9. The exception was the story recall evaluation for 1986, which was rated as comparatively worse for the unselected sample. With regard to postdictions, 9 of the 12 were equivalent. The unselected sample postdicted relatively lower scores for fact recall in 1995, 1992, and 1986. With regard to postdiction-performance difference scores, 11 of the 12 were similar. The exception was word

recall in 1995, which was greater for the unselected sample. In general, the pattern of missing data was as follows: For the unselected sample, a greater time lag was associated with a higher number of missing postdictions.

With regard to background indicators, it was shown that participants in the unselected sample were on average older by 2.8 years, less educated by 1.2 years, and less likely to be married by 18.9% than participants in the selected sample. This is a pattern similar to selective attrition effects in longitudinal research. This pattern also parallels one found by Hertzog et al. (1990b) in their study examining young, middle-aged and elderly people, who were assigned to either a prediction or no-prediction condition regarding performance levels on word and story recall. Similar to the present study, some of the portion of the sample in the prediction condition did not give performance estimates. It was found that participants omitting predictions were more likely to be elderly, had less years of education, and had lower performance scores than did the participants giving complete predictions. In addition, a Prediction x Age interaction for word recall showed that there were larger age differences among participants in the no-prediction condition than in the prediction condition. Although this may have indicated a differential impact of predictions on performance (e.g., older persons are more motivated after having made predictions), it may also have demonstrated that exclusion due to missing data positively biased the older prediction group because many less cognitively able persons omitted predictions. In the no-prediction condition, a parallel group of less cognitively able persons would presumably have existed, and the inclusion of their data would have decreased the

group's average performance level, thus accounting for the interaction. In the present study, a prediction-as-motivator mechanism could not have acted differently between groups to provide a potential explanation of the association of missing postdictions with slightly poorer performances. This suggests that through generating a pocket of missing data, the slightly less cognitively able may have deleted themselves from inclusion in the analyses.

The overall pattern of differences between the 2 samples suggests that a self-selection process occurred due to the likelihood of answering questions pertaining to memory beliefs. A sizeable group of lower performing, older, less educated and less likely married participants were less likely to speculate upon their memory performances than were the group of higher performing, younger, more educated and more likely married participants who were used in the analyses of the present study. Although these differences are significant, their magnitude is not severe. Nevertheless, this systematic omission of participants from the selected sample presents a dilemma of generalizability when interpreting the results. One cannot determine whether the same patterns found in the selected sample would have also held had the unselected participants provided sufficient data to be included in the analyses. However, given that the differences between the samples are not extreme, it is considered probable that the patterns observed in the selected sample would have held for the unselected group. Additionally, since both groups of participants were equally accurate in their performance assessments, it is most likely that they were employing the same method by which to generate specific assessments of past memory performance. If this is so,

the inferences drawn regarding the use of implicit theories of memory by the elderly should apply to the unselected sample equally.

An additional consideration is whether the present results would hold for the population of elderly in general. As noted earlier, the initial sample were recruited on the basis of acceptable levels of health, vision, hearing, and writing ability, and were self-selected on indices of education and occupation level. These differences are compounded by selective attrition over the four occasions of testing and the previously noted self-selection due to the completeness of data for the current occasion. These selection considerations combine to make the results of the present study more likely to apply to the portion of older adults in the general population who are similar to the present sample on the health, sensorimotor, demographic and socioeconomic indices. Although this group constitutes a large portion of the healthy older adult population, it is not representative of the entire population.

#### Importance of the Present Study

The concern over memory aging is a highly salient topic among elderly people, so related issues will become increasingly important as our society's demographics shift towards an "older" population. Metamemory questions addressed in the present study include the profile, accuracy of, and influences on, memory change beliefs.

These beliefs are important to the daily lives of aging adults as their exaggeration can lead to an unfounded fear over great memory loss that is anticipated or interpreted based on past experience. The widespread dissemination of information about the exaggerated nature of memory loss beliefs could help curb this worry for

many people. The message for policy makers and applied gerontologists is that there is some memory decline, but that the magnitude of decline is not as great as often believed. Furthermore, the shape of the decline function may vary by individual and memory task. Not all individuals experience the same degree of "memory aging" and not all memory tasks tap the same social and cognitive aging processes. Finally, this study reveals that older adults are generally accurate in their self-assessments of the profile of memory aging. Practitioners in gerontology may value such information.

In addition, beliefs about memory change, whether accurate or not, may affect the extent to which compensatory behaviours are adopted (Bäckman & Dixon, 1992). Compensatory behaviours may be performed by the self through such activities as making lists or rehearsing people's names, or by others through such behaviours as reminding a person of his/her appointments, or even structuring his/her life as is the case when memory deficits require institutionalization. The beliefs held about memory change also have the potential to affect the psychological adjustment of elderly people through fostering feelings of competence or incompetence. These feelings may, in turn, affect the welfare of the elderly by either prompting or disinclining them to elicit help from clinicians and social service providers as compensatory agents.

The appropriateness of help-seeking behaviour therefore depends upon the veridicality of these feelings, and this in turn affects the welfare of our society. If elderly people subscribing to false beliefs in memory decline come to feel incompetent in their lives and seek out help, the unnecessary burden placed upon social services

would be detrimental to the society at large. Conversely, if incompetent elderly people subscribing to false beliefs in memory maintenance fail to seek out appropriate help, the quality, and perhaps even safety, of their lives could suffer greatly. Perhaps equally important for the provision of appropriate care are clinicians and social services workers, who would greatly benefit from knowing if reports of memory change can be held to be reliable. A greater understanding of what elderly people believe to be true about their memory, the accuracy of these beliefs, and the influences upon these beliefs stands to benefit us all.

### Conclusion

Memory aging beliefs, their accuracy, and the influences upon them are becoming increasingly salient concerns as the age of populations increases. This study has addressed various questions related to these issues.

First, the profile of memory beliefs in the elderly was investigated. Results showed that the elderly generally believed that their performance on various memory tests had declined over 9 years, and was inferior to various comparative targets. These results support the idea that a theory of memory change, specifically memory decline, was operating.

Second, the accuracy of elderly persons' memory beliefs was examined. Performance on each task was found to be systematically overestimated on each occasion; however, the magnitude of decline overestimation was negligible. While the elderly may start from an erroneously conceived current reference point when estimating their past status, they are somehow able to generate a fairly good estimate

of the magnitude of decline experienced in their memory abilities over 9 years.

When accuracy was assessed for differences across occasions, word and fact recall showed greater accuracy for the current occasion than for retrospective ones, whereas story recall showed equal accuracy across occasions. For word and fact recall, the greater inaccuracy of past performance estimates was believed to reflect a greater ambiguity in recalling estimated performance levels. It was questioned whether performance would have been estimated spontaneously at the time of task completion and stored in memory over such a long period. Considered more feasible was the explanation that performance estimates were generated retrospectively, using the currently estimated status and a theory of memory aging as guides.

Third, memory self-efficacy was tested for its relationship with memory belief variables. MSE was found to be moderately associated with the perception of memory maintenance in a few variables. The relationship of memory belief variables to MSE was thought to be attenuated by task experience. In addition, it was suggested that certain memory tasks were more likely related to global MSE than were others.

Fourth, the process whereby postdictions for past occasions were generated was investigated. Results showed that current memory performance beliefs combined with memory change beliefs to predict the 2 most distant postdictions for both word and story recall, but not for fact recall. Overall, it was considered that different tasks may be more or less subject to theories of memory change, that it may be helpful to divide memory change scales into questions pertaining to similar aspects of memory, and that different tasks may be subject to different time lapses before a theory of memory

change is applied. Future research is necessary in order to clarify whether these considerations are valid.

The final area investigated was the existence of systematic differences between the selected and unselected samples. The unselected sample were found to differ on some key demographic and performance variables. It was suggested that the unselected sample were less cognitively able than the selected sample. Unfortunately, the sample differences present limits when attempting to generalize the results of the study to the unselected sample and the population in general. The two samples were equally accurate in their postdictions, which was considered unlikely if they were each employing different strategies by which to generate these assessments of memory performance. Therefore, the inferences drawn concerning the use of implicit theories of memory aging were considered to also apply to the unselected sample.

Overall, much support was found for an implicit theory explanation for the recall of memory capabilities. The study provided support for many elements of the hypothesis, and in turn generated more questions worthy of investigation. This research on the profile, accuracy of and influences upon memory change beliefs, contributes to our understanding of these metamemory questions, which in turn stands to benefit clinicians, social service providers, the elderly themselves, and society in general.

## References

- Bäckman, L. & Dixon, R. A. (1992). Psychological compensation: A theoretical framework. Psychological Bulletin, *112*, 259-283.
- Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. Developmental Psychology, *23*, 611-626.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. Developmental Psychology, *25*, 729-735.
- Bahrick, H. P. (1984). Replicative, constructive, and reconstructive aspects of memory: Implications for human and animal research. Invited Symposium Midwestern Psychological Association Meetings: Long-term memories: How durable, and how enduring? Physiological Psychology, *12*, 53-58.
- Battig, W. F., & Montague, W. E. (1969). Category names for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monographs, *80*, 1-46.
- Berry, J. M., & West, R. L. (1993). Cognitive self-efficacy in relation to personal mastery and goal setting across the life span. International Journal of Behavioral Development, *16*, 351-379.
- Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. Psychology and Aging, *3*, 249-257.
- Cavanaugh, J. C. (1996). Memory self-efficacy as a moderator of memory change. In F. Blanchard-Fields & T. M. Hess, (Eds.), Perspectives on cognitive

change in adulthood and aging (pp. 488-507). New York: McGraw-Hill.

Crockett, W. H., & Hummert, M. L. (1987). Perceptions of aging and the elderly. Annual Review of Gerontology and Geriatrics, 7, 217-241.

Devolder, P. A. (1993). Adult age differences in monitoring of practical problem-solving performance. Experimental Aging Research, 19, 129-146.

Devolder, P. A., Brigham, M C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. Psychology and Aging, 5, 291-303.

Dixon, R. A., (1989). Questionnaire research on metamemory and aging. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), Everyday cognition in adulthood and late life (pp. 394-415). New York: Cambridge University Press.

Dixon, R. A., Hertzog, C., & Hultsch, D. F. (1986). The multiple relationships among Metamemory in Adulthood (MIA) scales and cognitive abilities in adulthood. Human Learning, 5, 165-177.

Dixon, R. A., & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. Journal of Gerontology, 38, 682-688.

Dixon, R. A., Hultsch, D. F., & Hertzog, C. (1989). A manual of 25 three-tiered structurally equivalent texts for use in aging research (2nd ed., CRGCA Tech. Rep. No. 2). Victoria, British Columbia, Canada, University of Victoria, Department of Psychology.

Dixon, R. A., Hultsch, D. F., & Hertzog, C. (1988). The metamemory in adulthood (MIA) questionnaire. Psychopharmacology Bulletin, 24, 671-688.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual

for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Heckhausen, J, Dixon, R. A., & Baltes, P. B. (1989). Gains and losses in development throughout adulthood as perceived by different adult age groups. Developmental Psychology, 25, 109-121.

Hertzog, C., & Dixon, R. A. (1994). Metacognitive development in adulthood and old age. In J. Metcalfe & A.P. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 227-251). Cambridge, MA: MIT Press.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990a). Metamemory in adulthood: Differentiating knowledge, belief, and behavior. In Thomas M. Hess (Ed.), Aging and Cognition: Knowledge Organization and Utilization (pp. 161-212). New York: North-Holland.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990b). Relationships between metamemory, memory predictions, and memory task performance in adults. Psychology and Aging, 5, 215-227.

Hertzog, C., Hultsch, D. F., & Dixon, R. A., (1989). Evidence for the convergent validity of two self-report metamemory questionnaires. Developmental Psychology, 25, 687-700.

Hertzog, C., Saylor, L. L., Fleece, A. M., & Dixon, R. A. (1994). Metamemory and aging: Relations between predicted, actual and perceived memory task performance. Aging and Cognition, 1, 203-237.

Howard, D. V. (1980). Category norms: A comparison of the Battig and

Montague (1969) norms with the responses of adults between the ages of 20 and 80. Journal of Gerontology, 35, 225-231.

Hultsch, D. F., Hertzog, C., Dixon, R. A., & Small, B. J. (in press). Memory change in the aged. New York: Cambridge University Press.

Hultsch, D. F., Hertzog, C., Small, B. J., McDonald-Miszczak, L., & Dixon, R. A., (1992). Short term longitudinal change in cognitive performance in later life. Psychology and Aging, 7, 571-584.

Johansson, B., Allen-Burge, R., & Zarit, S. (1997). Self-Reports on Memory Functioning in a Longitudinal Study of the Oldest Old: Relation to Current, Prospective, and Retrospective Performance. Journal of Gerontology, 52, 139-146.

Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

McDonald-Miszczak, L., Hertzog, C., & Hultsch, D. F. (1995). Stability and accuracy of metamemory in adulthood and aging: A longitudinal analysis. Psychology and Aging, 10, 553-564.

McDonald-Miszczak, L., Hunter, M. A., & Hultsch, D. F. (1994). Adult age differences in predicting memory performance: The effects of normative information and task experience. Canadian Journal of Experimental Psychology, 48, 95-118.

McDougall, G. (1994). Predictors of metamemory in older adults. Nursing Research, 43, 212-218.

McFarland, C., Ross, M., & Giltrow, M. (1992). Biased recollections in older adults: The role of implicit theories of aging. Journal of Personality and Social

Psychology, 62, 837-850.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. Journal of Verbal Learning and Verbal Behaviour, 19, 338-368.

Ross, M., (1989). Relation of implicit theories to the construction of personal histories. Psychological Review, 96, 341-357.

Ryan, E. B., & Kwong See, S. (1993). Age-based beliefs about memory changes for self and others across adulthood. Journal of Gerontology, 48, 199-201.

Schaie, K. W., Willis, S. L., & O'Hanlon, A. M. (1994). Perceived intellectual performance change over seven years. Journal of Gerontology: Psychological Sciences, 49, 108-118.

Small, B. J., Dixon, R. A., Hultsch, D. F., & Hertzog, C. F. (1998). Longitudinal changes in quantitative and qualitative indicators of word and story recall in young-old and old-old adults. Manuscript submitted for publication.

Taylor, J. L., Miller, T. P., & Tinklenberg, J. R. (1992). Correlates of memory decline: A 4-year longitudinal study of older adults with memory complaints. Psychology and Aging, 7, 185-193.

Thurstone, L. L., & Thurstone, T. G. (1949). Examiner manual for the SRA Primary Mental Abilities Test. Chicago: Science Research Associates.

Zelinski, E. M., Gilewski, M. J., & Schaie, K. W. (1993). Individual differences in cross-sectional and 3-year longitudinal memory performance across the adult life span. Psychology and Aging, 8, 176-186.

Table 1

Comparison of Educational and Occupational Characteristics of the VLS Sample  
(Age 55-86) with the General Population of British Columbia and Canada in 1986

Years of Education	VLS Sample	British Columbia	Canada
1-4 years	0%	3%	7%
5-8 years	7%	24%	33%
9-10 years	11%	19%	19%
11-13 years	38%	33%	25%
14-17 years	35%	14%	11%
18+ years	9%	4%	3%
<b>Occupational Category</b>			
Professional and Semiprofessional	54%	27%	24%
Skilled	35%	22%	28%
Semiskilled	4%	30%	13%
Unskilled	6%	9%	31%
Other	0%	11%	4%

Table 2

## Characteristics of Cohort 1 of the VLS Sample, 1986 - 1995

Wave	1	2	3	4
Year of Testing	(1986)	(1989)	(1992)	(1995)
Total Sample Size	484	330	243	173
Number of Females	288	198	142	104
Number of Males	196	132	101	69
Reasons for Loss of Participants from the Sample				
Died		8	10	17
Health problems		23	23	21
Memory problems		5	2	0
Family health problems		10	2	1
Busy, not interested		83	29	22
Moved		13	8	4
Could not be located		3	7	3

Retrospective Evaluations' Means, Standard Deviations and T-Test Values of Means  
Compared to Value Signifying Status is "About the Same" as Present

	<u>M</u>	<u>SD</u>	t: Compared to a value of 2
<u>Word Recall</u>			
1986 compared to 1995	1.59	.92	4.52***
1989 compared to 1995	1.59	.71	5.85***
1992 compared to 1995	1.83	.59	2.95**
<u>Story Recall</u>			
1986 compared to 1995	1.79	.73	2.94**
1989 compared to 1995	1.75	.55	4.60***
1992 compared to 1995	1.82	.49	3.74***
<u>Fact Recall</u>			
1986 compared to 1995	1.58	.74	5.76***
1989 compared to 1995	1.67	.64	5.29***
1992 compared to 1995	1.76	.59	4.13***

\*Significant at  $p < .025$     \*\*Significant at  $p < .01$     \*\*\*Significant at  $p < .001$

Note: Scale ranged from 0 to 4, with 2 indicating "about the same", and lower values denoting a lower opinion of current performance compared to that in the past. T-tests were run on the means compared to a value of 2,  $df = 105$ .

Means and Standard Deviations for Numerical Postdictions, expressed in both  
Absolute and Percent Values

Occasion	<u>M</u>	<u>SD</u>	<u>M(%)</u>	<u>SD(%)</u>
<u>Word Recall</u>				
1986	20.45	4.93	68.16	16.43
1989	20.43	4.37	68.11	14.58
1992	19.48	4.54	64.94	15.13
1995	19.28	4.59	64.26	15.30
<u>Story Recall</u>				
1986	61.60	24.10	61.60	24.10
1989	61.01	23.70	61.01	23.70
1992	59.30	23.69	59.30	23.69
1995	56.96	23.95	56.96	23.95
<u>Fact Recall</u>				
1986	28.44	7.00	71.09	17.50
1989	27.47	7.12	68.68	17.79
1992	26.86	7.12	67.15	17.81
1995	26.09	7.21	65.22	18.03

Table 5

Relative Postdiction Means, Standard Deviations and T-Test Values of Means Compared to Value Signifying Status is

"About the Same" as Comparative Target

	<u>M</u>	<u>SD</u>	Rank Order	<u>t</u> : Compared to a Value of 2
<u>Word Recall</u>				
Myself compared to average peer	2.07	.67	1	.63
Myself compared to the best I could possibly do	1.65	.94	2	3.84***
Myself now compared to myself at age 20	1.51	1.05	3	4.80***
Myself compared to average 20-year-old	1.45	.96	4	5.88***
<u>Story Recall</u>				
Myself compared to average peer	2.29	.63	1	3.18**
Myself compared to the best I could possibly do	1.71	.82	2	3.69**
Myself now compared to myself at age 20	1.70	.98	3	3.18**
Myself compared to average 20-year-old	1.65	.87	4	4.11***

(table continues)<sub>oo</sub>

Table 5

Relative Postdiction Means, Standard Deviations and T-Test Values of Means Compared to Value Signifying Status is

"About the Same" as Comparative Target

	<u>M</u>	<u>SD</u>	Rank Order	t:	Compared to a Value of 2
<u>Fact Recall</u>					
Myself compared to average peer	1.96	.62	1	.63	
Myself compared to the best I could possibly do	1.90	.95	2	1.13	
Myself now compared to myself at age 20	1.75	1.03	3	2.45*	
Myself compared to average 20-year-old	1.59	1.01	4	4.13***	

Note: Scale ranged from 0 to 4, with 2 indicating "about the same", and higher values denoting a better opinion of the self compared to the target. T-tests were run on the means compared to a value of 2, df = 105.

\*Significant at  $p < .025$     \*\*Significant at  $p < .01$     \*\*\*Significant at  $p < .001$

Table 6

Significant Task x Target Contrasts for Targeted Comparisons

Tasks Contrasted	Targets Contrasted	F	$\eta^2$
fact/word recall	average peer/best I could possibly do	8.21**	.08
fact/word recall	average peer/myself at age 20	5.17*	.05
fact/word recall	average peer/average 20-yr-old	7.65**	.07
fact/story recall	average peer/best I could possibly do	22.88***	.18
fact/story recall	average peer/myself at age 20	8.48**	.08
fact/story recall	average peer/average 20-yr-old	8.98**	.08

\*Significant at  $p < .025$ \*\*Significant at  $p < .01$ \*\*\*Significant at  $p < .001$

Pearson Product Moment Correlations of Numerical Postdictions with Actual Performances and the Significance of the Differences Between Past and Present

Accuracy

Occasion	$\bar{r}$	Difference From				
		1986 <sup>a</sup>	1989 <sup>a</sup>	1992 <sup>a</sup>	1995 <sup>a</sup>	1995 <sup>b</sup>
<u>Word Recall</u>						
1986 <sup>a</sup>	.46***	--		*	***	***
1989 <sup>a</sup>	.43***	--		**	***	***
1992 <sup>a</sup>	.69***	*	**		*	***
1995 <sup>a</sup>	.83***	***	***	*		
1995 <sup>b</sup>	.87***	***	***	***		
<u>Story Recall</u>						
1986 <sup>a</sup>	.41***	--	--	--	--	--
1989 <sup>a</sup>	.48***	--	--	--	--	--
1992 <sup>a</sup>	.45***	--	--	--	--	--
1995 <sup>a</sup>	.50***	--	--	--	--	--
1995 <sup>b</sup>	.44***	--	--	--	--	--

(table continues)

Pearson Product Moment Correlations of Numerical Postdictions with Actual Performances and the Significance of the Differences Between Past and Present Accuracy

Occasion	$\underline{r}$	Difference From				
		1986 <sup>a</sup>	1989 <sup>a</sup>	1992 <sup>a</sup>	1995 <sup>a</sup>	1995 <sup>b</sup>
	<u>Fact Recall</u>					
1986 <sup>a</sup>	.58***		--	--	*	*
1989 <sup>a</sup>	.53***	--		--	**	**
1992 <sup>a</sup>	.57***	--			*	*
1995 <sup>a</sup>	.74***	*	**	*		
1995 <sup>b</sup>	.74***	*	**	*		

<sup>a</sup>Recall measured by average of 2 test versions at each occasion

<sup>b</sup>Recall on test immediately preceding the postdiction

--Not Significant    \*Significant at  $p < .05$     \*\*Significant at  $p < .01$     \*\*\*Significant at  $p < .001$

## Percent Overestimation for Numerical Postdictions

Occasion	Word Recall		Story Recall		Fact Recall	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
1986 <sup>a</sup>	4.80	15.37	26.05	22.03	15.98	14.58
1989 <sup>a</sup>	8.03	15.57	25.66	20.79	15.60	16.05
1992 <sup>a</sup>	3.40	10.97	22.15	21.22	15.00	15.33
1995 <sup>a</sup>	3.40	8.57	21.46	21.04	14.58	12.30
1995 <sup>b</sup>	2.20	7.77	21.37	21.52	14.40	12.13

Note. All means were derived by subtracting the performance score from the numerical postdiction. Means and standard deviations were converted to percentages for ease of comparison across tasks.

<sup>a</sup>Recall measured by mean of 2 tests at each occasion

<sup>b</sup>Recall on test immediately preceding the postdiction

Table 9  
Age, Occasion and MSE Differences in Postdictions as Revealed in Hierarchical Multiple Regressions

Predictive Variable	Block in which Value Reported	t	Degrees of Freedom	R <sup>2</sup>	β
<u>Word Recall</u>					
Overall Model	1	3.49***	3, 420	.080	
Age	1	-5.28***			-.26
Occasion	1	-2.30*			-.11
<u>Story Recall</u>					
Overall Model	2	2.14***	6, 417	.062 (ΔR <sup>2</sup> = .052)	
Age	2	4.32***			-1.28
MSE	2	4.64***			3.15
Age x MSE	2	-4.74***			-3.04
<u>Fact Recall</u>					
Overall Model	1	2.25**	3, 420	.035	
Occasion	1	-2.51*			-.12
MSE	1	2.54*			.13

\*Significant at p<.05    \*\*Significant at p<.01    \*\*\*Significant at p<.001

Table 10

The Prediction of Past Numerical Postdiction using Current Numerical Postdiction, MIA Capacity and MIA Change

Task & Year	F: Overall Model	R <sup>2</sup> : Overall Model	f: 1995 Postdiction	Unique Variance in 1995 Postdiction	f: MIA Capacity	Unique Variance in MIA Capacity	f: MIA Change	Unique Variance in MIA Change
<u>Word Recall</u> (3, 102 degrees of freedom in overall model)								
1986	12.95***	.276	5.07***	.173	2.37*	.000	-3.80***	.103
1989	21.05***	.382	7.52***	.339	2.07	.003	-2.58*	.040
1992	33.23***	.494	9.76***	.480	1.66	.008	-1.08	.006
<u>Story Recall</u> (3, 102 degrees of freedom in overall model)								
1986	108.22***	.761	17.98***	.738	.77	.003	-2.93**	.020
1989	175.04***	.837	22.89***	.828	.901	.000	-.234*	.009
1992	262.89***	.885	27.92***	.881	1.88	.001	-1.77	.004

(table continues)

Table 10

The Prediction of Past Numerical Postdiction using Current Numerical Postdiction, MIA Capacity and MIA Change

Task & Year	F:	R <sup>2</sup> :	F:	Unique Variance in 1995 Postdiction	F:	Unique Variance in MIA Capacity	F:	Unique Variance in MIA Change
<u>Fact Recall (3, 102 degrees of freedom in overall model)</u>								
1986	74.36***	.686	14.19***	.679	.75	.000	-1.55	.007
1989	82.11***	.707	15.08***	.706	.66	.000	-.57	.001
1992	103.67***	.753	16.94***	.751	.88	.002	-.33	.000

Note. Values labelled "Overall Model" were arrived at in the third step of a hierarchical multiple regression, entering 1995 postdiction in step 1, MIA Capacity in step 2, and MIA Change in step 3, using the postdiction for the past occasion as the dependent variable. All  $F$  values listed are those attained for the overall model. "Unique Variance" for each variable is the  $R^2$  change for the step in which the variable was added.

\*Significant at  $p < .05$     \*\*Significant at  $p < .01$     \*\*\*Significant at  $p < .001$

Table 11

Means, Standard Deviations and t-tests for Mean Differences between Selected and Unselected Samples

Variable	Selected Sample <u>M</u>	Selected Sample <u>SD</u>	Unselected Sample <u>M</u>	Unselected Sample <u>SD</u>	<u>n</u>	t-value
<u>Subject Background Variables</u>						
Age	75.21	5.56	77.97	5.52	67	3.19**
Years of Education	14.86	2.92	13.63	3.25	67	2.59*
Marital Status	.54	.92	.97	1.11	67	2.66**
<u>Performance Variables</u>						
word recall, 1995 <sup>a</sup>	18.26	4.06	16.01	5.20	67	3.02**
word recall, 1992	18.46	3.51	17.55	4.60	67	1.38
word recall, 1989	18.02	4.35	17.55	4.54	64	.67
word recall, 1986	19.01	3.70	18.19	4.31	67	1.34
story recall, 1995 <sup>a</sup>	35.50	8.56	30.58	11.63	67	2.99**

(table continues)

Table 11

Means, Standard Deviations and t-tests for Mean Differences between Selected and Unselected Samples

Variable	Selected Sample		Unselected Sample		n	t-value
	M	SD	M	SD		
story recall, 1992	37.15	9.39	34.00	9.85	67	2.11*
story recall, 1989	35.35	10.77	35.59	11.00	61	.14
story recall, 1986	35.55	9.33	32.73	10.82	67	1.82
fact recall, 1995*	20.20	5.77	17.75	6.40	67	2.61**
fact recall, 1992	20.86	5.94	18.41	5.69	67	2.68**
fact recall, 1989	21.23	5.86	19.74	5.51	64	1.64
fact recall, 1986	22.04	5.16	19.97	5.52	67	2.51*
<u>Metamemory Variables</u>						
MIA - achievement	3.61	.46	3.68	.42	66	1.06
MIA - anxiety	2.97	.66	3.14	.68	67	1.62
MIA - capacity	2.97	.52	3.01	.53	67	.51

(table continues)<sub>8</sub>

Table 11

Means, Standard Deviations and t-tests for Mean Differences between Selected and Unselected Samples

Variable	Selected Sample		Unselected Sample		n	t-value
	M	SD	M	SD		
MIA - change	2.57	.64	2.54	.68	67	.27
MIA - locus	3.35	.65	3.38	.60	66	.26
MIA - strategy	3.82	.45	3.62	.55	67	2.63*
MIA - task	4.01	.40	4.01	.29	66	.01
<u>Retrospective Numerical Postdictions</u>						
word postdiction, 1995	19.28	4.59	17.87	6.36	62	1.53
word postdiction, 1992	19.48	4.54	18.66	4.64	47	1.03
word postdiction, 1989	20.43	4.37	19.86	5.10	43	.69
word postdiction, 1986	20.45	4.93	20.61	5.44	41	.17
fact postdiction, 1995	26.09	7.21	23.07	7.71	58	2.5*

(table continues)

Table 11

Means, Standard Deviations and t-tests for Mean Differences between Selected and Unselected Samples

Variable	Selected Sample		Unselected Sample		n	t-value
	M	SD	M	SD		
fact postdiction, 1992	26.86	7.12	23.86	8.06	35	2.09*
fact postdiction, 1989	27.47	7.12	24.90	8.70	31	1.68
fact postdiction, 1986	28.44	7.00	25.13	8.62	30	2.16*
story postdiction, 1995	56.96	23.95	52.62	25.13	61	1.11
story postdiction, 1992	59.30	23.69	54.89	24.83	45	1.03
story postdiction, 1989	61.01	23.70	56.37	25.19	41	1.05
story postdiction, 1986	61.60	24.10	56.21	26.05	39	1.17

Note. Selected sample  $N = 106$ ; values include those imputed using sample mean. Unselected sample  $N = 67$ ;  $n$  with valid data is listed for each value.

\*Recall measured by average of 2 test versions at each occasion

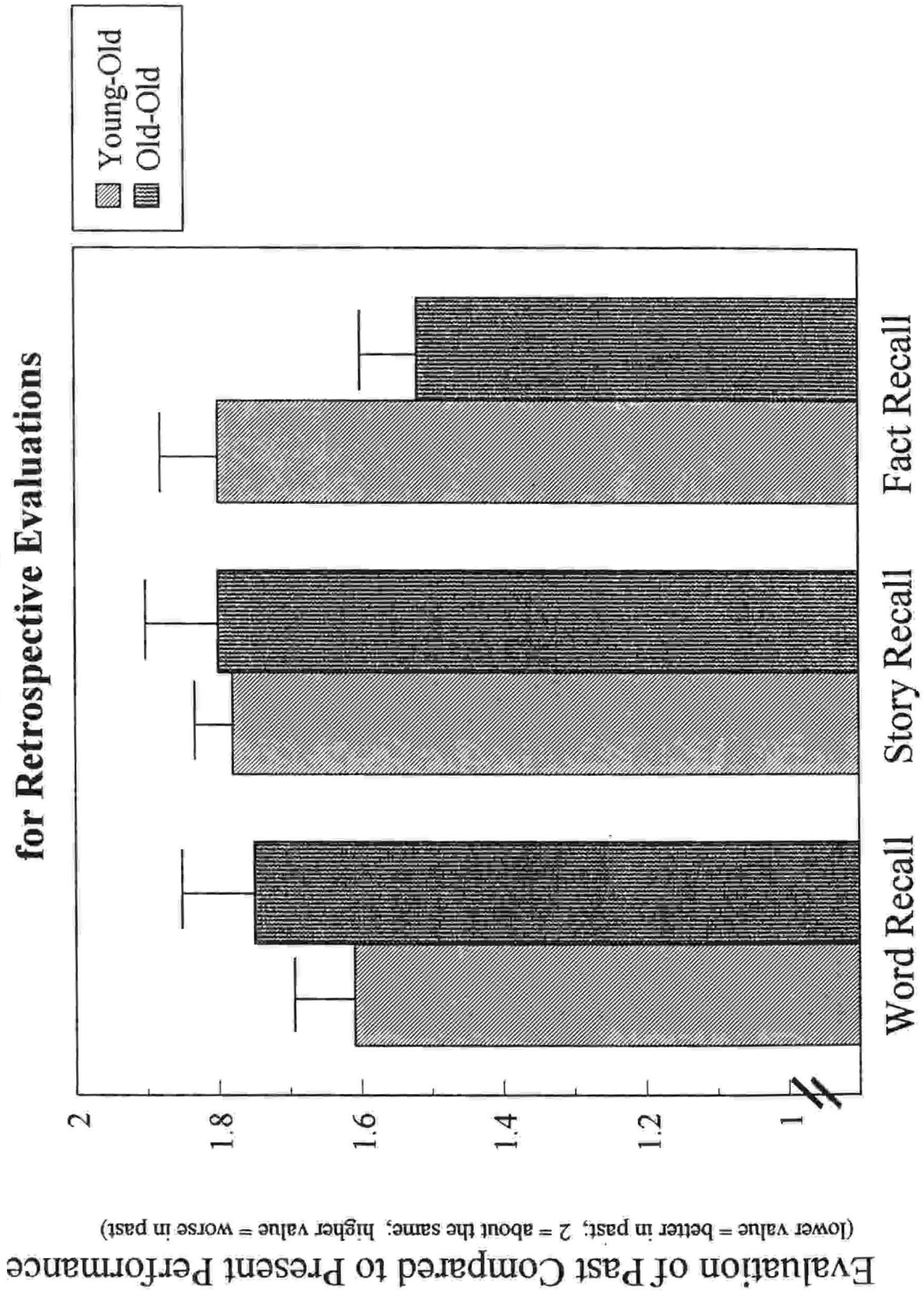
<sup>b</sup>Mean response per item within the subscale, ranging from 1 to 5

\*Significant at  $p < .05$  \*\*Significant at  $p < .01$  \*\*\*Significant at  $p < .001$

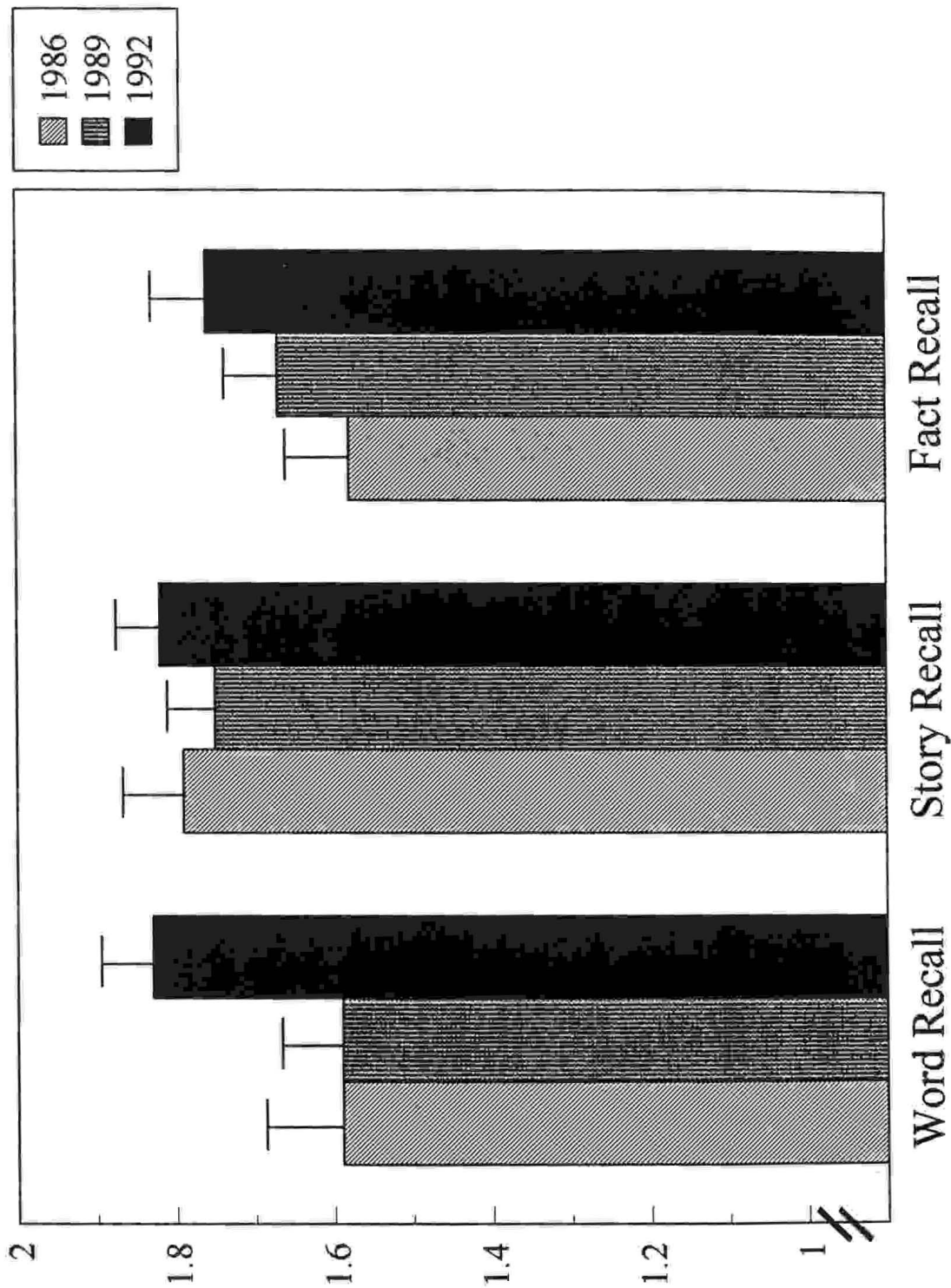
Breakdown of Educational and Marital Differences Between Selected and Unselected Samples

Variable & Classification	% of Selected Sample	% of Unselected Sample
<u>Education Level Achieved</u>		
No degree/diploma	8.5	17.9
High school diploma	35.8	37.3
Technical/community college certificate	17.0	19.4
Bachelor's degree	27.4	22.4
Master's degree	8.5	1.5
Any advanced degree (MD, PhD etc.)	2.8	1.5
Total with Bachelor's degree or higher	38.7	25.4
<u>Marital Status</u>		
Married	72.6	53.7
Single	3.8	3.0
Widowed	20.8	37.3
Divorced/separated	2.8	6.0
Total not married	27.4	46.3

**Figure 1**  
**Task x Age Group Interaction**  
**for Retrospective Evaluations**

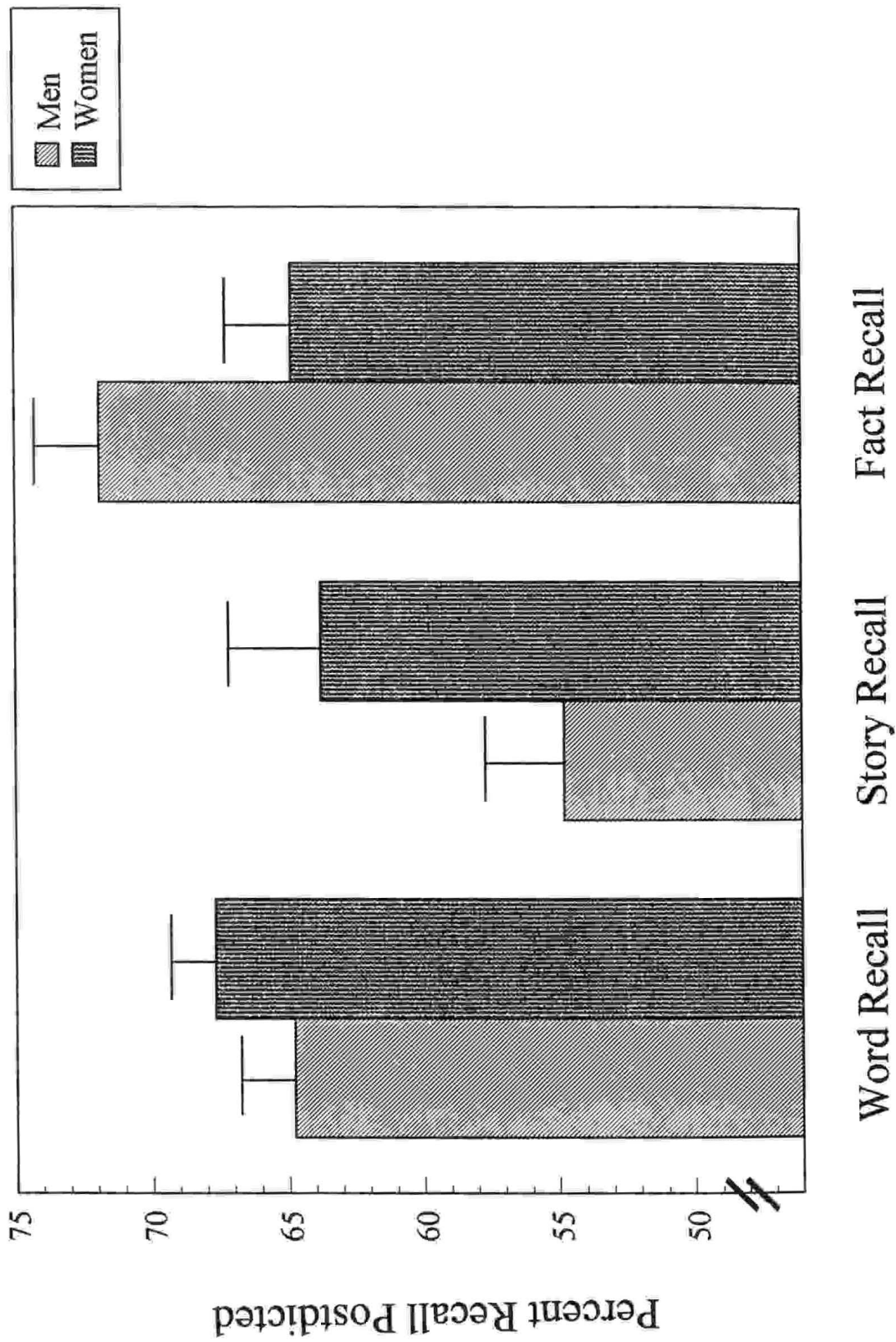


**Figure 2**  
**Task x Occasion Interaction**  
**for Retrospective Evaluations**

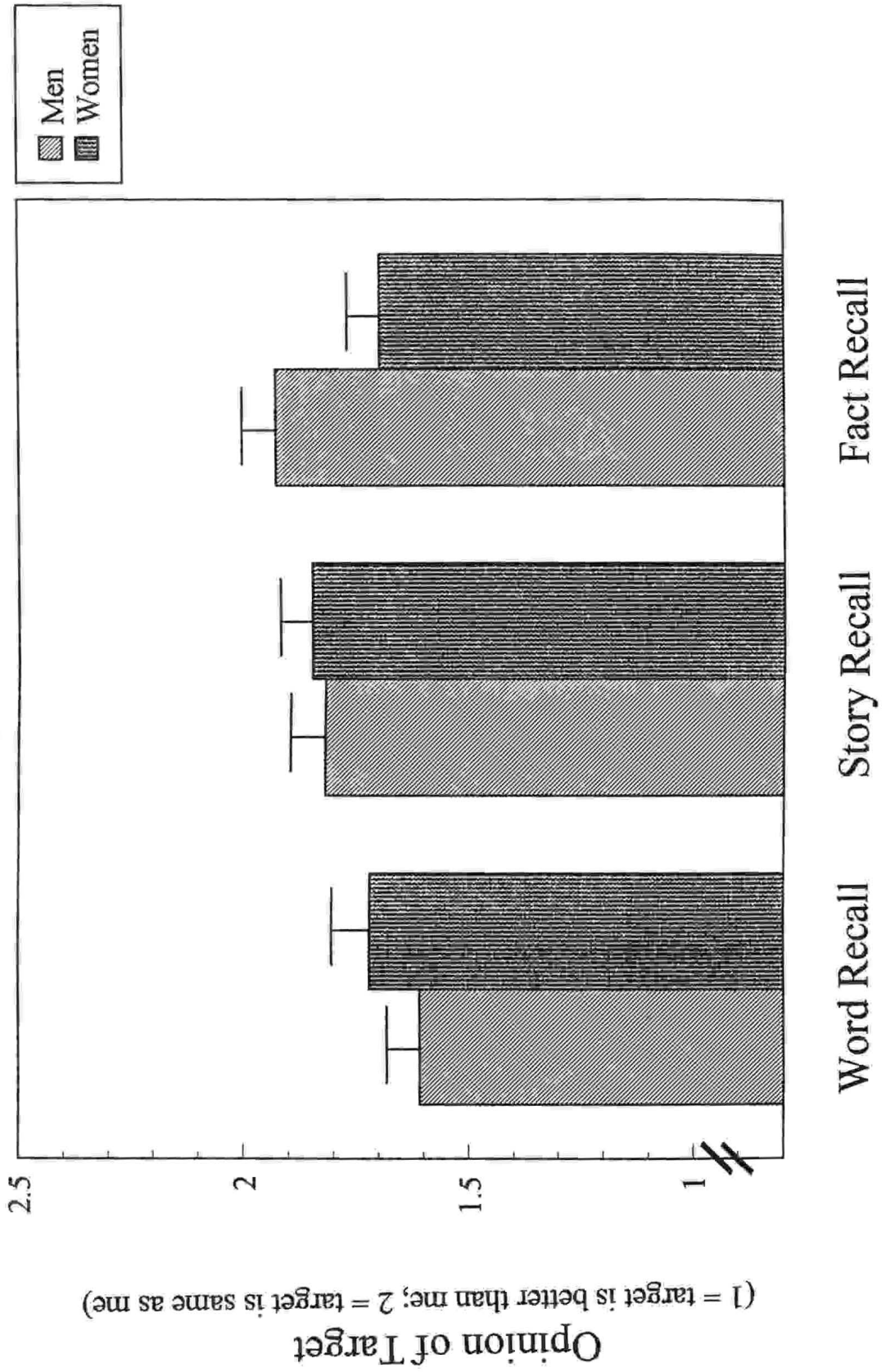


Evaluation of Past Compared to Present Performance  
 (lower value = better in past; 2 = about the same; higher value = worse in past)

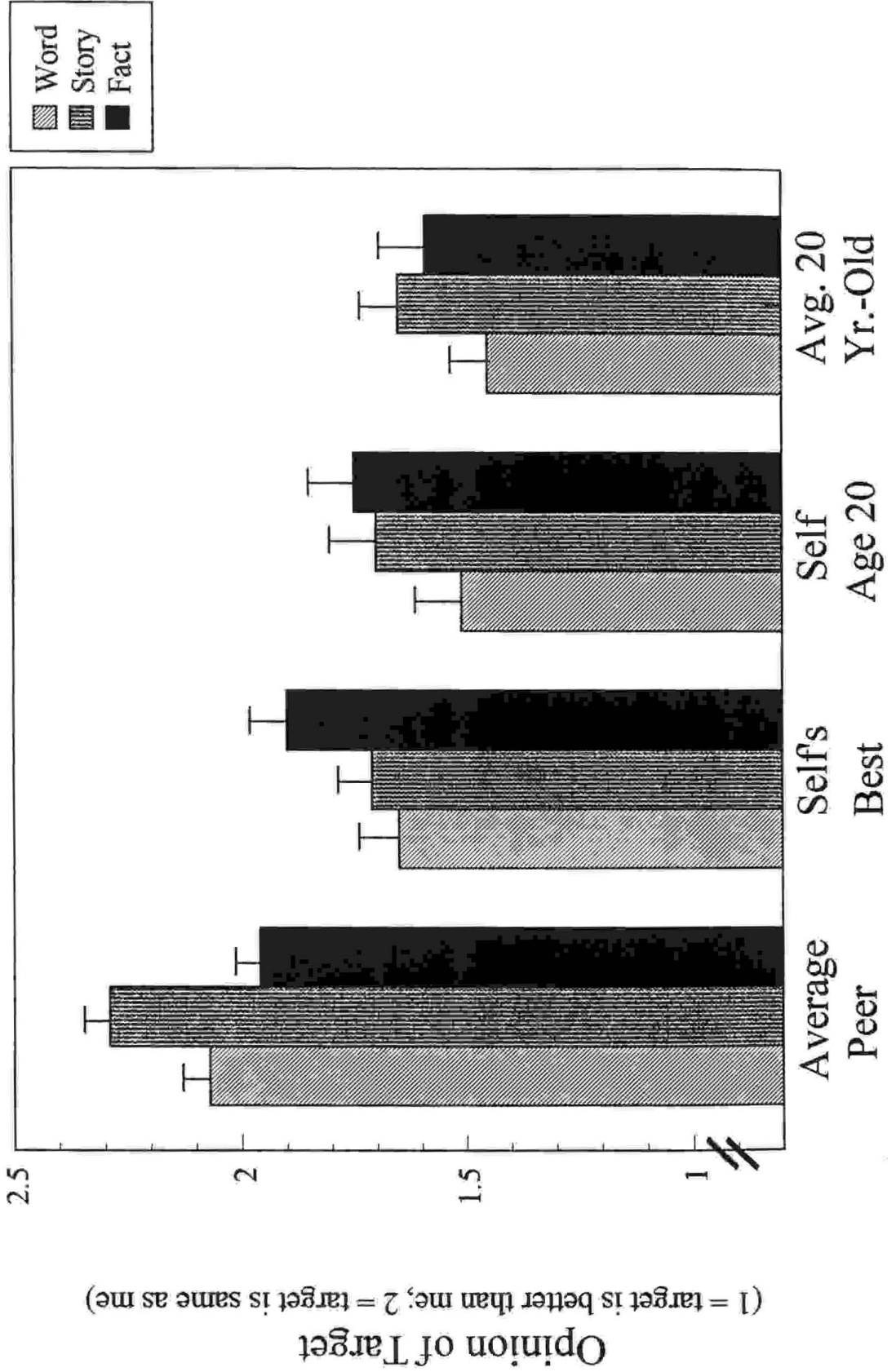
**Figure 3**  
**Task x Gender Interaction**  
**for Numerical Postdictions**



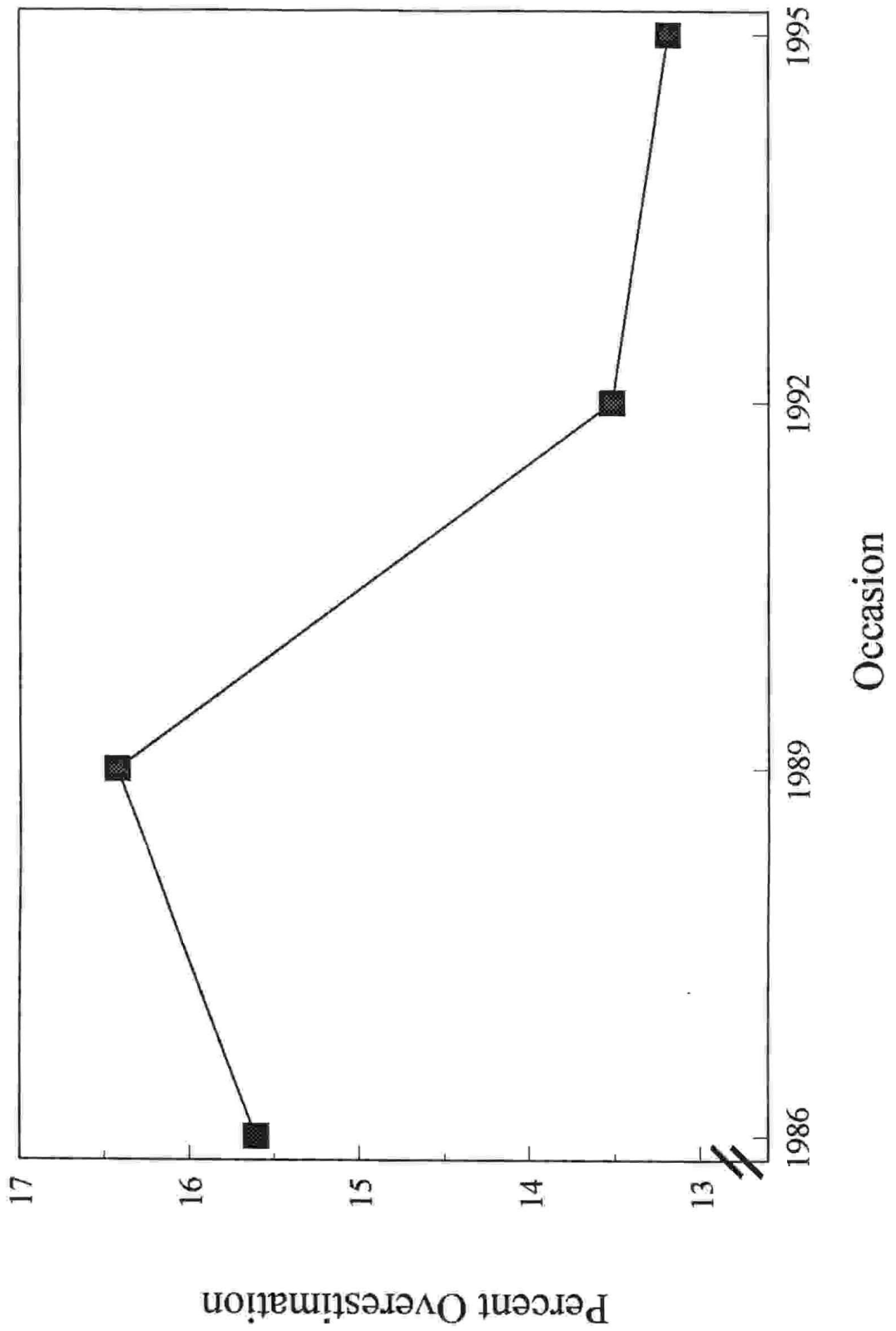
**Figure 4**  
**Task x Gender Interaction**  
**for Targeted Comparisons**



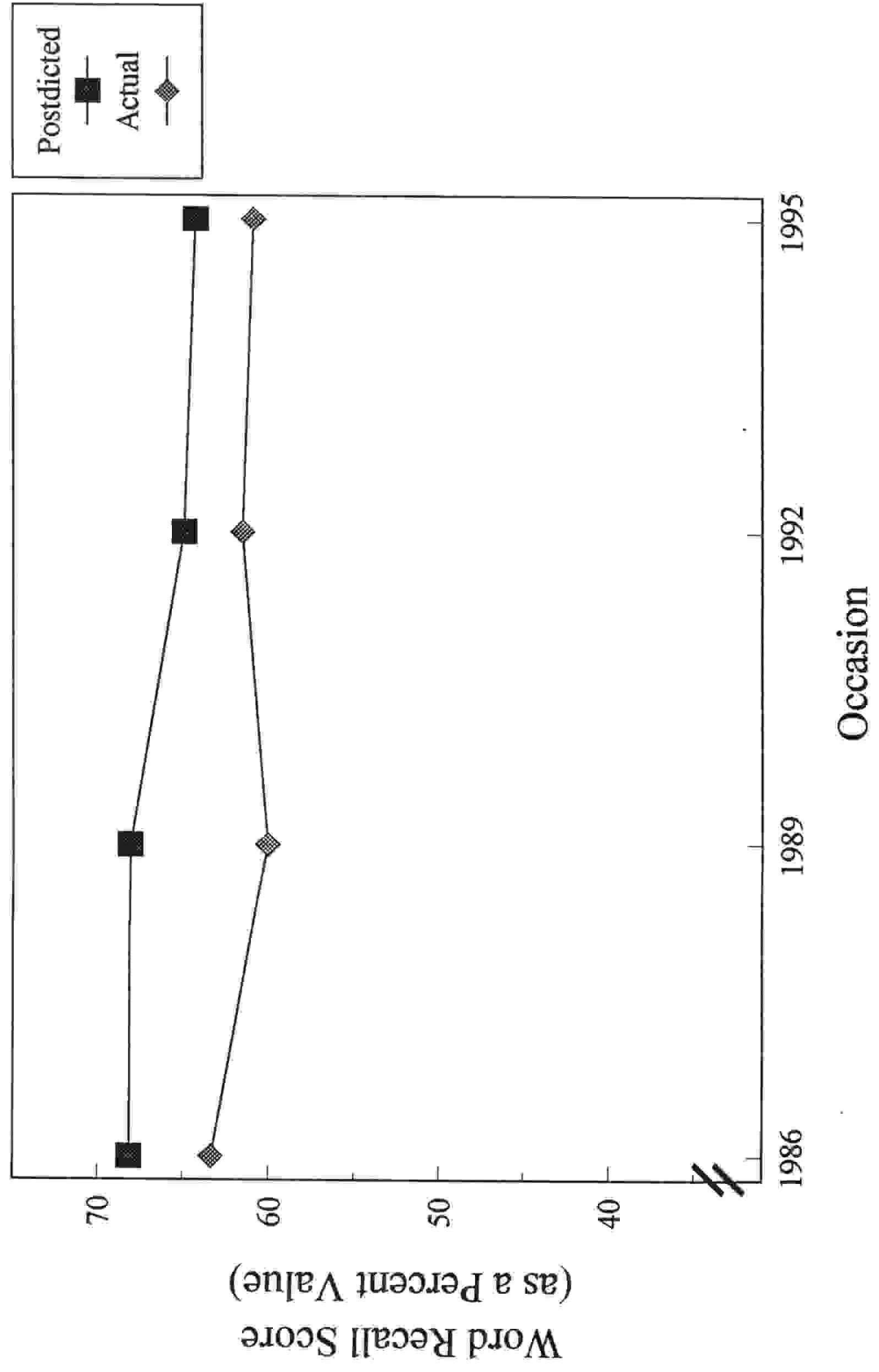
**Figure 5**  
**Task x Target Interaction**  
**for Targeted Comparisons**



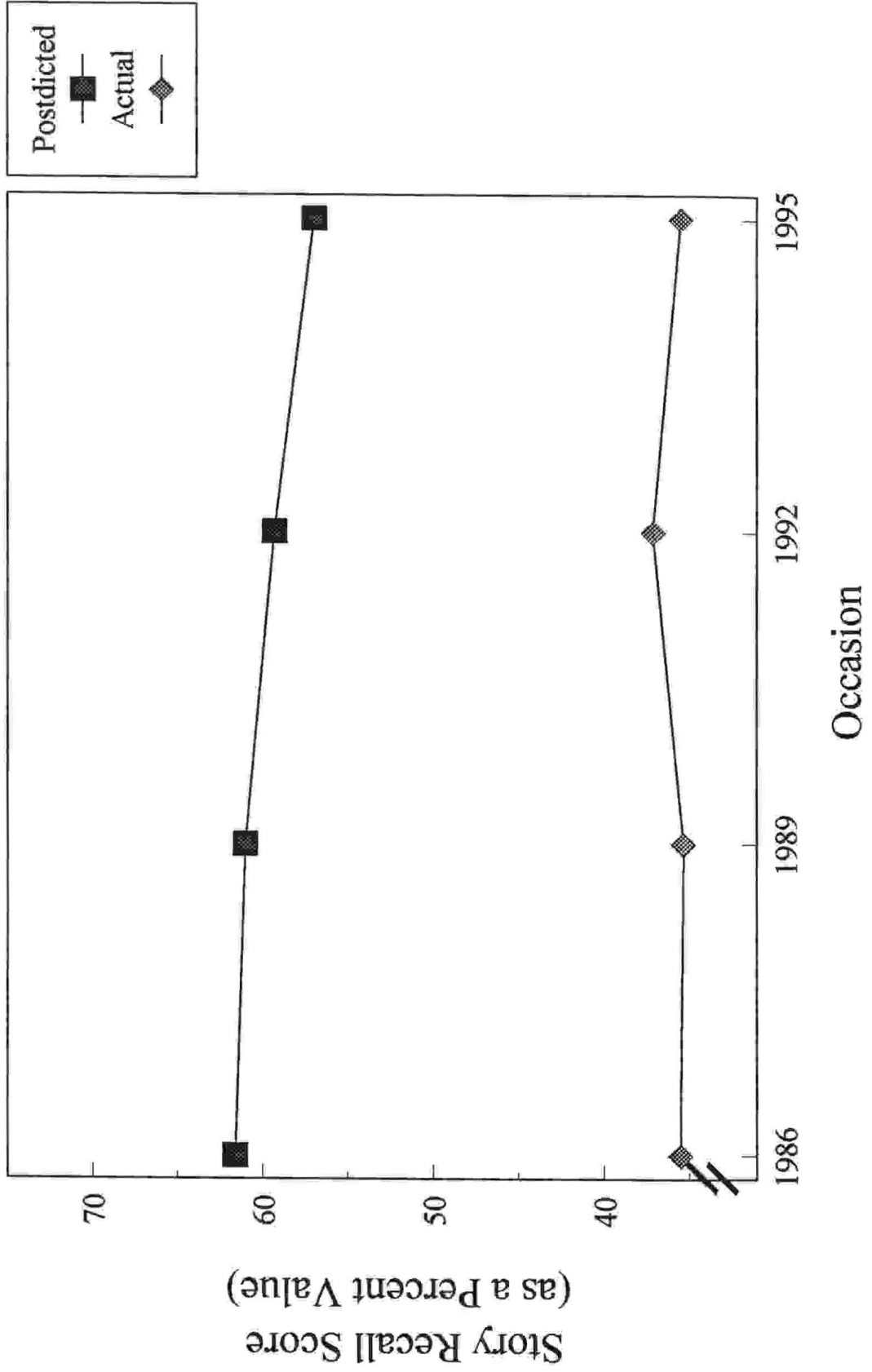
**Figure 6**  
**Occasion Effect of Accuracy**



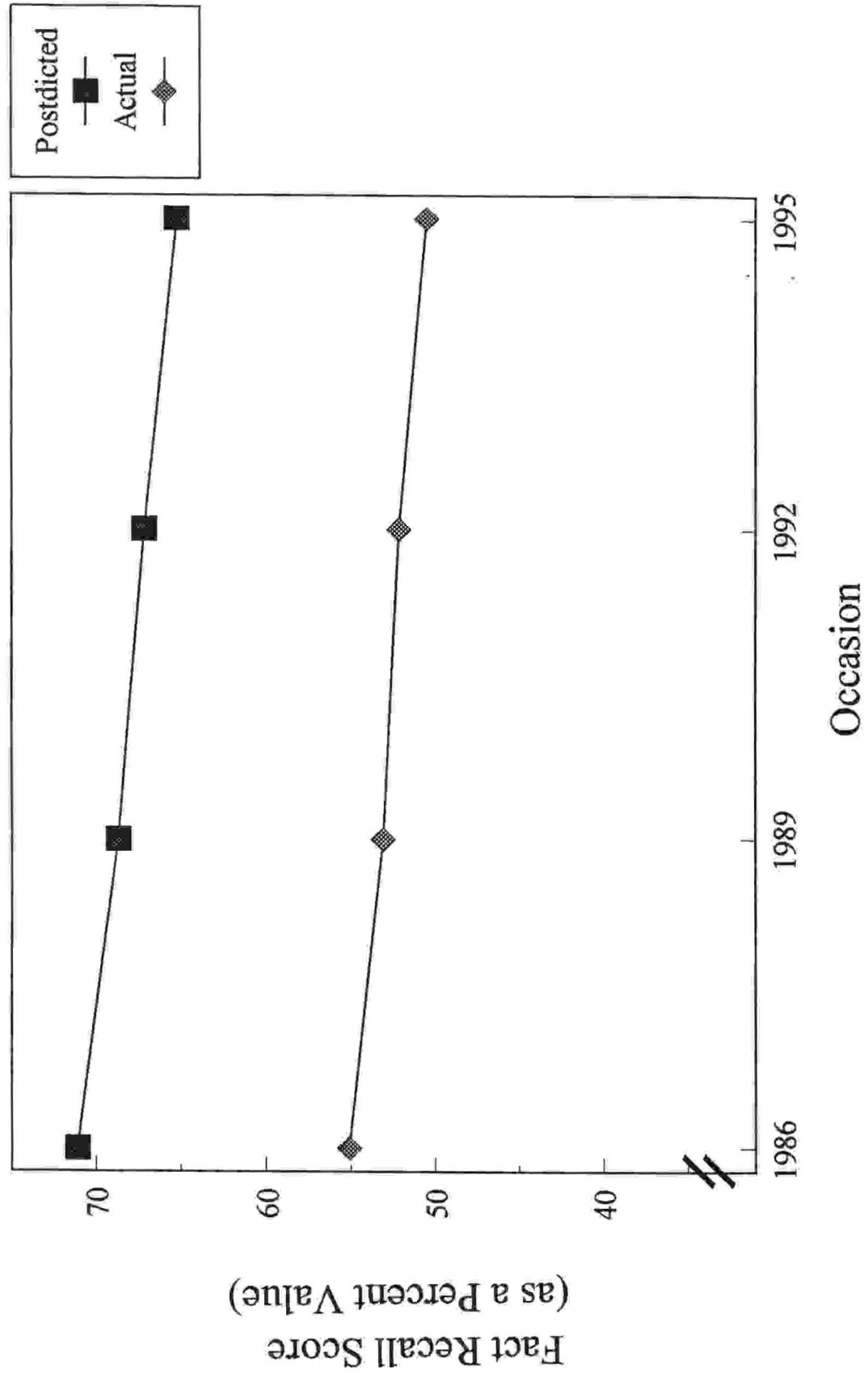
**Figure 7**  
**Postdicted and Actual Performance Scores (as Percentages)**  
**Over 4 Occasions for Word Recall**



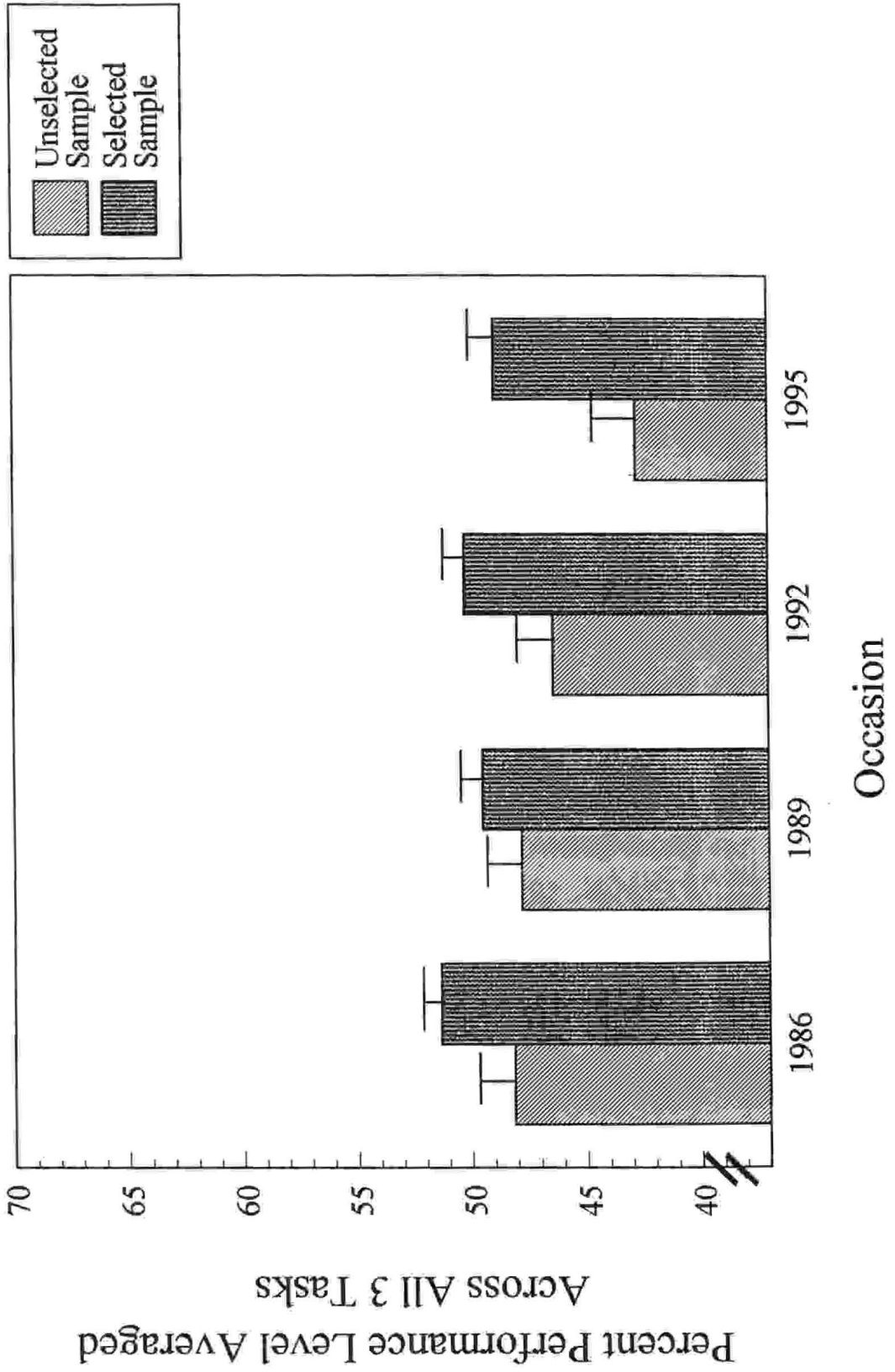
**Figure 8**  
**Postdicted and Actual Performance Scores (as Percentages)**  
**Over 4 Occasions for Story Recall**



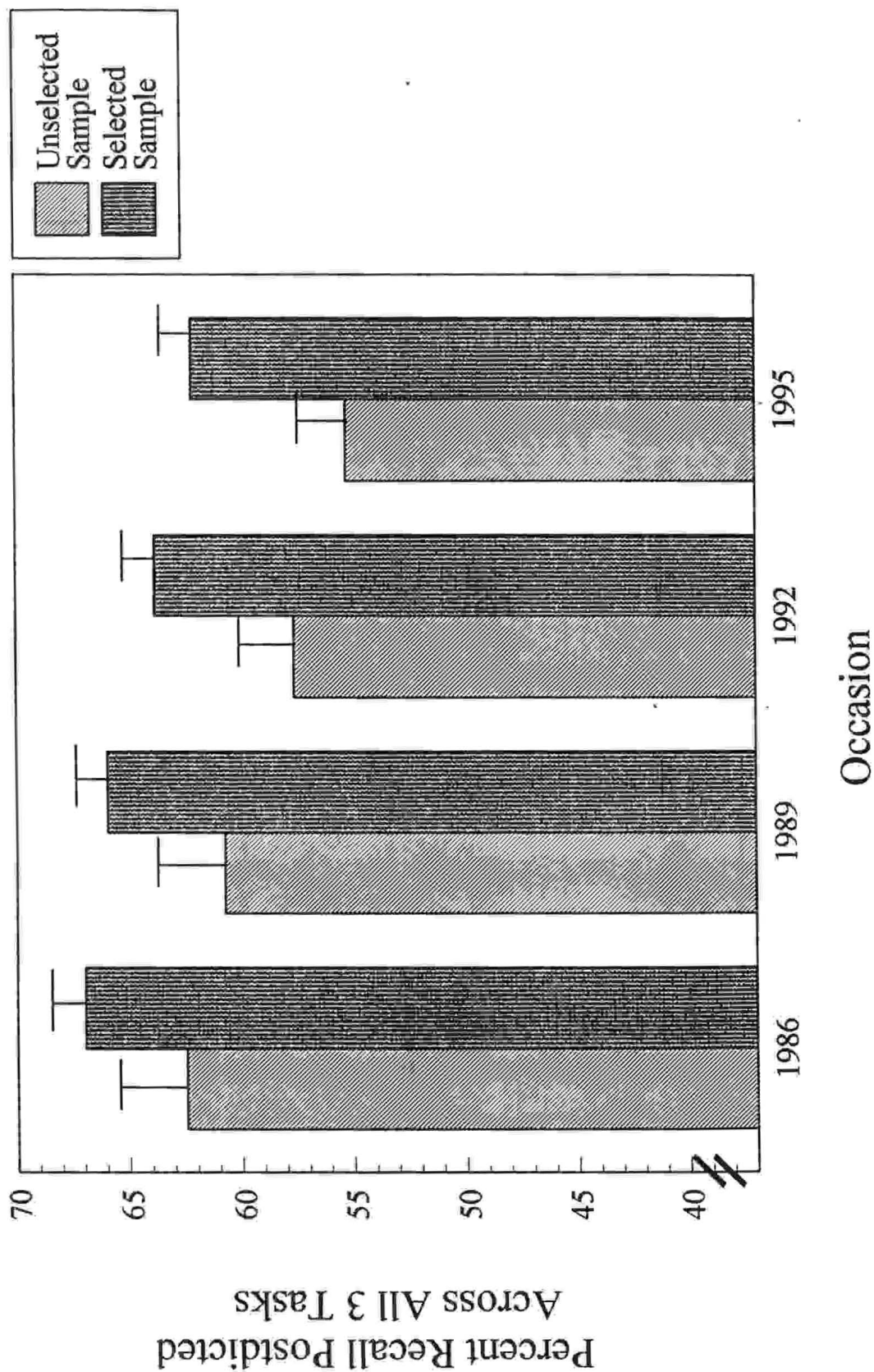
**Figure 9**  
**Postdicted and Actual Performance Scores (as Percentages)**  
**Over 4 Occasions for Fact Recall**



**Figure 10**  
**Sample x Occasion Interaction**  
**in Performance Levels**



**Figure 11**  
**Sample Differences in Numerical Postdictions**  
**Across Occasions**



## Appendix A

The 17 Items of the MIA Capacity Subscale

## Item    Wording

3.     I am good at remembering names.
9.     I am good a remembering birthdates.
19.    I have no trouble keeping track of my appointments.
27.    I am poor at remembering trivia.
49.    I am good at remembering the order that events occurred.
52.    I am good at remembering conversations I have had.
59.    I often forget who was with me at events I have attended.
62.    I am good at remembering the places I have been.
71.    I have no trouble remembering where I have put things.
77.    I am good at remembering things like recipes.
88.    I am good at remembering titles of books, films, or plays.
91.    I have no trouble remembering lyrics of songs.
95.    I am good at remembering names of musical selections.
97.    After I have read a book I have no difficulty remembering factual information from it.
100.   I am good at remembering the content of news articles and broadcasts.
104.   Remembering the plots of stories and novels is easy for me.
105.   I am usually able to remember exactly where I read or heard a specific thing.

## Appendix B

The 18 Items of the MIA Change Subscale

- | Item | Wording   |
|------|---|
| 10.  | I can remember things as well as always.  |
| 14.  | I'm less efficient at remembering things now than I used to be.   |
| 16.  | The older I get the harder it is to remember clearly.   |
| 18.  | I am just as good at remembering as I ever was.   |
| 28.  | I am much worse now at remembering the content of news articles and broadcasts than I was 10 years ago. |
| 30.  | Compared to 10 years ago, I am much worse at remembering titles of books, films, or plays.              |
| 32.  | I remember my dreams much less now than 10 years ago.   |
| 38.  | I misplace things more frequently now than when I was younger.  |
| 39.  | As people get older they tend to forget where they put things more frequently.                          |
| 41.  | Compared to 10 years ago, I now forget many more appointments.  |
| 45.  | My memory for important events has improved over the last 10 years.                                     |
| 54.  | My memory for phone numbers will decline as I get older.  |
| 56.  | My memory for dates has declined greatly in the last 10 years.  |
| 58.  | My memory for names has declined greatly in the last 10 years.  |
| 76.  | I know of someone in my family whose memory improved significantly in old age.                          |
| 82.  | My memory has improved greatly in the last 10 years.  |

## Appendix B (Continued)

The 18 Items of the MIA Change Subscale

Item	Wording
------	---------

89.	My memory has declined greatly in the last 10 years.
-----	--

92.	My memory will get better as I get older.
-----	---

## Appendix C

Wording of the Retrospective Evaluation**HOW DID YOU DO ON THE WORD MEMORY TASK?**

**What was the task?:** In this task, we asked you to study a list of 30 words for 2 minutes. The words were common English nouns representing different categories of things like birds, flowers, furniture, occupations, and musical instruments. After you studied the words, we asked you to recall and write down as many of them as you could in any order.

Do you recognize this task as one you did during the last few weeks?

\_\_\_\_\_ Yes \_\_\_\_\_ No

Now we would like you to tell us how you think you did on this task on each of the three previous occasions you gave participated in the VLS.

How did you do on this task 3 years ago (1992) compared to this time (1995)?

- a. much better 3 years ago
- b. somewhat better 3 years ago
- c. about the same
- d. somewhat worse 3 years ago
- e. much worse 3 years ago

## Appendix C (Continued)

Wording of the Retrospective Evaluation

How did you do on this task 6 years ago (1989) compared to this time (1995)?

- a. much better 6 years ago
- b. somewhat better 6 years ago
- c. about the same
- d. somewhat worse 6 years ago
- e. much worse 6 years ago

How did you do on this task 9 years ago (1986) compared to this time (1995)?

- a. much better 9 years ago
- b. somewhat better 9 years ago
- c. about the same
- d. somewhat worse 9 years ago
- e. much worse 9 years ago

Note. Each of the preceding questions was also asked in relation to story recall and fact recall.

## Appendix D

Wording of the Numerical Postdiction Questionnaire

Now we would like you to think about how you did on the task you just completed. There were 30 words on the list you were asked to remember. How many of the 30 words do you think you recalled correctly?

You were asked to remember similar lists of 30 words on the three previous occasions (1992, 1989, 1986) in which you participated in the VLS. We would like you to reflect on how you did on this memory task on each of these previous occasions.

How many of the 30 words do you think you recalled correctly 3 years ago (1992)?

How many of the 30 words do you think you recalled correctly 6 years ago (1989)?

How many of the 30 words do you think you recalled correctly 9 years ago (1986)?

## Appendix E

Wording of the Targeted Comparison Questionnaire

Now we would like you to think again about how you did on the word recall task you just completed during this session. This time we would like you to tell us how you think you did relative to what other people may have done. You just tried to recall the list of 30 words.

How do you think you did relative to the average person your age?

How do you think you did relative to the average 20-year-old?

How do you think you did relative to how you would have done when you were 20 years of age?

How do you think you did relative to the best you could possibly do (e.g., after practice or training)?

## VITA

Surname: Crow

Given Names: Carolyn Barnes

Place of Birth: Whitefield, Greater Manchester, England

### Educational Institutions Attended:

University of Calgary 1983 to 1988

University of Victoria 1995 to 1998

### Degrees Awarded:

B.A. University of Calgary 1988

### Honours and Awards:

University of Calgary Undergraduate Merit Award 1986

### Publications:

Dixon, R. A., Gagnon, L. M. & Crow, C. B. (in press). Collaborative memory accuracy and distortion: Performance and beliefs. In M. J. Intons-Peterson & D. Best (Eds.), Memory distortions and their prevention. Mahwah, N.J.: Lawrence Erlbaum Associates Inc.

PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Title of Thesis:

"The Way We Were": Retrospective Evaluations of Memory Performance in the Elderly.

Author



Carolyn Barnes Crow

April 30, 1998