Was that Part of the Story or Did I Just Think So?
Age Differences, Mild Cognitive Impairment, and Intraindividual Variability in Inferences and Story Recognition

by

Allison Anne Marie Bielak
B. A., University of Winnipeg, 2002

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCES

in the Department of Psychology

© Allison Anne Marie Bielak, 2004
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.
The present study expanded the story recognition and inference literature by investigating age differences within the older age range, differences as a result of mild cognitive impairment (MCI), and extending the focus of the investigation into the consistency of responding. 304 older adults completed a story recognition task across five different occasions. Old-old (OO) adults and those with more severe MCI showed poorer ability to accurately recognize inferences, and less sensitivity to discriminate between statement types. Intraindividual variability was positively correlated with increasing age and cognitive impairment, and interactions revealed the greatest inconsistency involved the false, rather than inferred statements. The findings support our proposal that participants used two different recognition strategies, and their episodic memory ability defined the efficiency and frequency of use of the strategies. OO and MCI adults may be less able to recognize that something plausible and consistent with an event may not have actually occurred.
# Table of Contents

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page .........................................................</td>
</tr>
<tr>
<td>Abstract .........................................................</td>
</tr>
<tr>
<td>Table of Contents ...............................................</td>
</tr>
<tr>
<td>List of Tables ...............................................</td>
</tr>
<tr>
<td>List of Figures ...............................................</td>
</tr>
<tr>
<td>Acknowledgements ..........................................</td>
</tr>
<tr>
<td>Introduction ...............................................</td>
</tr>
<tr>
<td>Cognitive Aging and Specificity in Episodic Memory</td>
</tr>
<tr>
<td>Story Structure ...........................................</td>
</tr>
<tr>
<td>Qualitative Age Differences in Story Recall: Main Idea versus Specific Item Recall</td>
</tr>
<tr>
<td>Story Comprehension and Inferences ......................</td>
</tr>
<tr>
<td>Age Differences in Inferring ..................................</td>
</tr>
<tr>
<td>Memory Differences within the Aged .......................</td>
</tr>
<tr>
<td>Story Memory and Alzheimer's Dementia ...................</td>
</tr>
<tr>
<td>Story Memory and Mild Cognitive Impairment ............</td>
</tr>
<tr>
<td>Intraindividual Variability ..................................</td>
</tr>
<tr>
<td>Greater Intraindividual Variability with Age ............</td>
</tr>
<tr>
<td>Intraindividual Variability and Older Clinical Populations</td>
</tr>
<tr>
<td>Intraindividual Variability and Mild Cognitive Impairment</td>
</tr>
<tr>
<td>Task Differences in Intraindividual Variability ........</td>
</tr>
<tr>
<td>Present Study ...............................................</td>
</tr>
<tr>
<td>Research Questions ...........................................</td>
</tr>
<tr>
<td>1. Are there Differences within Old Age in the Ability to Recognize Inferences?</td>
</tr>
<tr>
<td>2. Does Cognitive Status Play a Role in the Ability to Recognize Inferences?</td>
</tr>
<tr>
<td>3. Are there Differences within Old Age in the Amount of Inconsistency in Distinguishing Inferences?</td>
</tr>
<tr>
<td>4. Is Cognitive Status a Significant Factor in the Amount of Intraindividual Variability in Recognizing Inferences?</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Education, Health and Cognitive Variables as a Function of Age and Cognitive Status ................................................................. 40

Table 2: Mean Accuracy Scores to Each Statement Type by Age and Cognitive Status ........................................................................... 50

Table 3: Mean Latency Scores to Each Statement Type by Age and Cognitive Status ........................................................................... 69

Table 4: Correlations of Intraindividual Variability with Overall Accuracy, Signal Detection Measures, and Mean Latency ......................... 83
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean Overall Accuracy by Age and Statement Type</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>Mean Overall Accuracy by Cognitive Status and Statement Type</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>Mean $d'$ (True/False) as a Function of Age and Cognitive Status</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>Mean $d'$ (True/Inference) as a Function of Age and Cognitive Status</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>Mean $d'$ (True/False) and $d'$ (True/Inference) as a Function of Age</td>
<td>63</td>
</tr>
<tr>
<td>6</td>
<td>Mean $d'$ (True/False) and $d'$ (True/Inference) as a Function of Cognitive Status</td>
<td>64</td>
</tr>
<tr>
<td>7</td>
<td>Difference Between $d'$ (True/False) and $d'$ (True/Inference) by Age and Cognitive Status</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>Mean criterion (True/False) and criterion (True/Inference) by Age</td>
<td>67</td>
</tr>
<tr>
<td>9</td>
<td>Mean criterion (True/False) and criterion (True/Inference) by Cognitive Status</td>
<td>68</td>
</tr>
<tr>
<td>10</td>
<td>Mean Latency as a Function of Statement Type and Age</td>
<td>71</td>
</tr>
<tr>
<td>11</td>
<td>Mean Latency as a Function of Statement Type and Cognitive Status</td>
<td>73</td>
</tr>
<tr>
<td>12</td>
<td>Mean Latency Residual $T$-scores for Each Participant as a Function of Age</td>
<td>76</td>
</tr>
<tr>
<td>13</td>
<td>Average ISD as a Function of Age and Cognitive Status</td>
<td>79</td>
</tr>
<tr>
<td>14</td>
<td>Average ISD as a Function of Age and Statement Type</td>
<td>80</td>
</tr>
<tr>
<td>15</td>
<td>Average ISD as a Function of Cognitive Status and Statement Type</td>
<td>81</td>
</tr>
</tbody>
</table>
Acknowledgements

The very first person I need to thank is my advisor, Dr. David Hultsch, for his willingness to answer my many questions, read through my many drafts of this thesis, and guide me in the right direction both scholastically and about the academic world in general over the past two years. I definitely made the right choice in coming to the University of Victoria to learn from him. I would like to thank my other committee members, Dr. Esther Strauss and Dr. Helena Kadlec, and also Dr. Mike Hunter, for offering their expertise and encouragement throughout this process. My other supports throughout these past two years have been invaluable and came from many sources. First, I need to thank the entire team at the Victoria Longitudinal Study: Aislin, Laura, Jackie, Debbie, Doug, Dianne, Veronique, Terry, and Stuart. My time spent over at the VLS was like a vacation from the stresses of grad school, and you were all always there to listen to my concerns and anxieties with a sympathetic ear, help to calm my fears, and offer support and encouragement every step of the way. Next, I must also thank my wonderful office mates, Ben and Tanya, who never sighed at hearing another of my many questions, helped me to work over hurdles in my course or thesis work, and always offered helpful advice either in the office or during one of our ritualistic afternoon coffee/brain breaks. Graduate school is such a new and different experience that being able to share this process with others has made it more bearable and definitely a lot more fun. Thanks Karen, Marei, Sue, Jing, and Marianne. Finally, I have to thank the people closest to me, my parents, who always offered unwavering support and optimism. I never would have made it without the two of you.
Introduction

Cognitive Aging and Specificity in Episodic Memory

Cognitive aging is not uniform. There is considerable variability between individuals in the rate of age-associated cognitive change, and significant differentiation across cognitive domains in the characteristics and speed of decline. Generally speaking, performance in fluid intellectual domains that involve processing speed and executive functions show relatively greater amounts of decline over time compared with changes in crystallized abilities such as vocabulary and general knowledge (Light & Burke, 1988; Park, 2000). For example, Schaie (1983) found that age-related decline on tasks measuring verbal meaning began later in the aging process than other abilities such as logical reasoning and visuo-spatial skills. Similarly, Cunningham, Clayton, and Overton (1975) demonstrated that while older adults had significantly lower scores on a test of fluid intelligence, the magnitude of this difference was considerably greater than age differences on a vocabulary test.

There is also differentiation within the area of memory function. Episodic memory, which involves the recall of event-specific information, shows substantially more age-related decline than the recall of general knowledge (semantic memory), or the retention of procedural skills (procedural memory; Craik, 2000). However, the majority of research involving episodic memory and aging has employed single unrelated words as the to-be-recalled stimuli. As a result, some investigators proposed that episodic memory for more meaningful materials (i.e., discourse) would be more resistant to the detrimental effects of cognitive aging (e.g., Hulicka, 1967). Texts may have an enhanced likelihood of recall because they have more meaning to an individual than a list of words (Hess &
Pullen, 1996; Hultsch, Hertzog, & Dixon, 1984; Zelinski & Gilewski, 1988), which may increase the motivational level of older adults (Nesselroade & Labouvie, 1985). In addition, because any communication with another person typically involves a description of past or current events, recall of connected discourse is a very familiar, everyday task (Zacks & Hasher, 1988). Finally, comprehension of a story requires the active integration of prior knowledge with the new incoming verbal information (Klatzky, 1988; Hertzog, Dixon, & Hultsch, 1992), and this additional processing may result in superior recall compared with a list of words. Consequently, text recall's high ecological validity may result in observing a smaller decline with age (Salthouse, 1991).

These factors support the possibility that age-related decline in story recall may look quite different from other recall tasks. Despite the unique attributes of connected discourse, younger adults routinely outperform older adults on story recall tasks (Adams, 1991; Adams, Labouvie-Vief, Hobart, & Dorosz, 1990; Cohen, 1979; Dixon, Hultsch, Simon, & von Eye, 1984; Dixon, Simon, Nowak, & Hultsch, 1982; Hartley, 1988; Hultsch, Hertzog, & Dixon, 1984, 1990; Hultsch, Masson, & Small, 1991; Spilich, 1983; Tun, 1989; Zelinski, Gilewski, & Thompson, 1980), as they do on word recall tasks (e.g., Hultsch et al., 1990, 1991). However, age differences in story and word recall are not identical; story recall does appear to show less decline over time than word recall. For example, Small, Dixon, Hultsch, and Hertzog (1999) found significant longitudinal decline in older adults' word recall ability after 6 years, but observed much less decline on a story recall task. Thus, although the existence of quantitative age differences in story recall reiterates the diminished memory capacity typically found in aging, the age
differences appear to be smaller than those seen for memory of less coherent material, such as a word list.

Due to the differences between word and story recall in age-related decline, it may be the case that story recall proceeds down a different path of decline than other episodic memory stimuli. There may not only be quantitative age differences in story recall, but qualitative age-related differences in the actual processing or comprehension of text material as well. More specifically, do older adults simply recall less material than younger adults, or do they actually recall different material? In order to clearly describe the research surrounding this question, it is necessary to first provide a brief summary of story structure.

**Story Structure**

According to Kintsch’s (1974) theory of text meaning, a story is made up of propositions, or arguments characterized by a verb which specifies the relationship among one or more arguments. These propositions are hierarchically arranged according to their similarity or relevance to a main idea in the story. A superordinate proposition represents a main idea in the story, whereas a proposition which expands on and provides further detail of that main idea is called a subordinate proposition. In other words, a superordinate proposition has a higher level in the story’s hierarchical structure, and subordinate propositions occupy the lower levels. In order to coherently understand a story it is necessary to cognitively string together the higher level propositions with the lower level propositions, thereby focusing on their interrelationships.

When reading or listening to a story, it is generally the case that an individual attempts to remember the more important points of the story (gist) or higher level items,
rather than focusing on recalling all the details of the story or lower level items. This is termed the levels effect, and refers to the tendency to recall items of higher levels of importance over those of lower levels which do not supply as much information (Meyer, 1975).

Both younger and older adults consistently adhere to the levels effect and recall a greater proportion of the major ideas of a story compared with the minor details (Adams, Smith, Nyquist, & Perlmutter, 1997; Dixon et al., 1982; Stine & Wingfield, 1990; Tun, 1989). This suggests that older adults maintain the ability to organize and discriminate the hierarchical structure of the text despite their overall lower quantity of text propositions recalled. However, there is also evidence that the type of information (i.e., gist or detail) proportionally recalled by older adults is different from younger adults.

**Qualitative Age Differences in Story Recall: Main Idea versus Specific Item Recall**

The majority of qualitative story recall research finds that older adults do not differ from younger adults in their proportionate recall of higher level propositions, but do recall significantly less lower level information (Spilich, 1983; Stine & Wingfield, 1987; Zelinski et al., 1980). The magnitude of this effect is dependent on a variety of text, reader, and task variables (Hultsch & Dixon, 1984; Meyer & Rice, 1989), such as text organization (Byrd, 1981), text type (Hess & Pullen, 1996), verbal ability (e.g., Dixon et al., 1984), and meaningfulness of material (Hultsch & Dixon, 1983).

Some researchers explain this qualitative age difference as a compensatory strategy to offset older adults’ declining working memory ability. They propose that in response to a general decline in processing resources and working memory, older adults reallocate these limited resources to focus primarily on the theme of the text rather than
the details. This strategy ensures that the primary ideas of the text are retained at the expense of less critically important details ("loss with compensation").

Stine and Wingfield (1990) concluded that older adults only demonstrate a decrease in their proportionate differentiation between the main ideas and the details of a story under demanding task conditions. This increase in processing requirements can be caused by either individual factors, such as verbal ability (e.g., Dixon et al., 1984, Stine & Wingfield, 1988), or because of contextual demands, such as the organization of text (e.g., Byrd, 1981). For example, Tun (1989) evaluated the hypothesis that narratives with highly predictable organizational schemes facilitate processing in older adults by providing built-in organization, thus requiring fewer cognitive resources. Older and younger adults recalled either a narrative text or an expository passage which has a more loosely defined organizational scheme. Both groups recalled significantly more propositions and demonstrated greater differentiation in the hierarchy of propositions (levels effect) for narrative recall versus expository recall. The supposition was that the demands of the narrative text on working memory were lower than that of the expository task. In addition, Meyer and Rice (1989) concluded that greater age deficits are found when the text is presented at a fast pace versus a slower presentation or self-paced reading. Furthermore, Hultsch, Hertzog, and Dixon (1990) found that age-related differences in text memory are primarily predicted by individual performance on both verbal speed and working memory tasks.

There is another body of research that has investigated qualitative age differences in propositional story recall from an entirely different perspective. This line of research is derived from a life-span developmental viewpoint, and questions the traditional
assumption that any cognitive changes that accompany aging are detrimental in nature. Researchers from this tradition propose that older adults' lesser recall of details of a story is a natural result of development ("loss due to growth"). Although younger adults may excel at detailed recall of surface-level information, older adults undergo a developmental shift towards a higher level integrative and informationally dense recall style. Older adults are uniquely able to interpret the significance of new information in light of their past experiences and knowledge, and are able to extract the moral or metaphoric meaning of narratives. They focus more on the psychological and symbolic processes than on the logical and analytic ones. Moreover, this hypothesis of developmental change with aging is consistent with the social perspective, whereby the older adult typically passes lessons and cultural traditions on to the younger cohorts. The "cost" of this developmental shift only appears as a loss because detail recall is compared to younger adults who are perhaps developmentally more suited to score higher on the tasks typically used to assess story recall (i.e., number of propositions recalled vs. explaining the moral of the story). This view suggests age-related deficits in story recall may reflect developmentally normal changes in thought structure.

Evidence supporting this hypothesis comes from Adams, Labouvie-Vief, Hobart, and Dorosz (1990), who found that given older adults' more interpretive response style, they automatically recalled the story's moral, whereas younger adults did not until explicitly asked to do so. Adams (1991) confirmed that narrative recall shows a significant developmental shift from text-based to interpretive responding across early and late adolescents, and middle-aged and older adults; the number of text-based recall units increased up until middle-age, at which point it dropped, but the production of the
story's metaphoric meanings continued to increase. Finally, Adams, Smith, Nyquist, and Perlmutter (1997) investigated the differences in older and younger adults' ability to retell and interpret a Sufi tale. Sufi tales have at least two levels of meaning; one is a relatively simple action-event sequence found in the story’s propositional content, and the other is rich in psychological content found in several levels of symbolic meanings. The participants' responses were rated on their depth and synthesis of interpretation, where a deeper interpretation drew away from the literal meaning towards a more symbolic one, and higher synthesis indicated the symbolic elements of the story were unified as a complete whole rather than explained as individual concepts (analytic). As predicted, the older adults produced more deep and synthetic interpretations than the younger adults, who produced more analytic (deep or shallow) responses.

Although the developmental shift hypothesis is definitely supported by these findings, Adams and her colleagues (1997) noted that either hypothesis of qualitative age-related changes in story recall could be correct. Both of these theories have support in the literature, and the accuracy of one over the other has not yet been demonstrated. Despite this, the fact that fewer details are recalled from a story with increasing age has implications for story comprehension.

**Story Comprehension and Inferences**

Stories are unique in the type of processing they require for comprehension. To illustrate, both expository and narrative texts are organized in a hierarchical fashion with ideas relating to a central topic. However, in an expository text (e.g., encyclopaedia entry), the connections among the ideas are specifically stated, whereas in narratives, the relations among the idea units are not all explicit and consequently some actions or
statements must be inferred. Inference refers to “the reasoning involved in making a logical judgment on the basis of circumstantial evidence and prior conclusions rather than on the basis of direct observation” (Dictionary.com, 2003). For example, read the sample story below (Dixon, Hultsch, & Hertzog, 1989):

Jane was excited about going to the beach with her parents. They had not gone to the shore since she was 9 years old, when they had visited Halifax. This time would be better, because they were going to stay for a week at Long Beach. Her parents had told her of the beautiful white sand, the clear water, and the miles of unspoiled beach with very few people. She was hoping, though, that she might see a few friends her own age.

While not explicitly stated, the reader might infer that Jane plans on taking her swimsuit with her, given the common knowledge that when attending a beach, one typically wears a swimsuit. It is assumed that the reader has some general world knowledge and is able to reason and fill in the gaps in the story by integrating their prior knowledge with the new information. Consequently, one’s ability to infer can be deemed a marker of comprehension (Hamm & Hasher, 1992).

Age Differences in Inferring

Some research has found that older adults show a reduced ability to draw inferences relative to younger adults. Among the first, Cohen (1979, Experiment 1) found that after reading a story, older adults were significantly worse at responding to questions that required inferences to be drawn. She even found that older adults made significantly more errors than younger adults in identifying anomalies in brief passages (e.g., a housewife who had run out of bread made sandwiches; 1979, Experiment 2), but in many cases the older adults were wrong because they responded with an unexpected (non-target) anomaly based on their own personal value judgements (e.g., she should
have a hot lunch not sandwiches). Cohen concluded that while inferring ability appears to decline with age, this deficit is likely due to processing overload rather than memory loss because the older adults experienced no difficulty in answering the questions that relied on verbatim memory (Experiment 1).

Despite this conclusion, Cohen (1981) re-evaluated the possibility that older adults are unable to construct inferences because they forget the recently presented facts. She noted that while verbatim recall requires only single items to be retained, making delayed inferences requires the simultaneous recall of several facts. Therefore, age differences in drawing inferences should disappear when the task is not dependent on memory recall because the needed information remains available for review. As expected, older adults made significantly more errors in judging the truth of a potential inference from a story when they listened compared to read the story. From this, Cohen proposed that the listen method might have exceeded the processing capacity of older adults, thus hampering their ability to construct inferences at the same time. Surprisingly, there were no significant differences for both the younger and older adults between the two memory conditions in the read condition (i.e., allowed to look back at the story or not), demonstrating that the difficulties in inferring that seem to accompany age are not due to forgetting the relevant information. Cohen's second experiment (1981) investigated age differences in inferences derived from factual knowledge (e.g., "A burning cigarette was carelessly discarded. The fire destroyed many acres of virgin forest."); requires the factual knowledge that burning cigarettes can start fires), but the results still led to the conclusion that difficulties in answering the inference questions could not be due to failures in recalling the target information. She instead concluded
that older adults maintain their ability to draw inferences, but are simply less efficient at doing so.

Although Light, Zelinski, and Moore (1982) also discovered age differences in inferring, they concluded that the decline had to be due to forgetting the relevant information. In their first experiment, older adults did not show any performance differences on the inference questions if they were forced to process the facts more fully by inferring during the task, given examples and a strategy on answering the questions, or given no special instructions at all. Therefore, the older adults maintained their ability to draw inferences, and the poorer performance of the older adults compared to the younger adults on both the fact and inference questions given after the task suggested the involvement of memory resources rather than reasoning ability. The failure to find a significant difference in the older adults’ performance when they were also given a context for the three sentence set to encourage deeper-level processing (1982, Experiment 2) again demonstrated that lower processing resources were not the cause of age differences in inferring. Light et al.’s final experiment (1982) varied the order of the facts from which an inference had to be drawn. For example, a linear ordering of the argument, AB, BC, and CD, is easy to follow because the points follow from one another in a logical manner. In contrast, CD, AB, and BC, is far more difficult, because CD and AB have nothing in common with one another and must be stored until BC is presented, and even then, the elements must be rearranged to form a cohesive argument. Light et al. proposed that if older adults have less working memory capacity, they should have more difficulty inferring statements from out of order arguments than from those in a linear arrangement. In fact, older adults did have proportionately more difficulty in correctly
inferring when the statements were in orders that were more demanding of working memory. Light and her colleagues consequently discounted the possibilities that older adults have an impaired ability to reason, do not understand the task, or inadequately process the incoming information. They instead concluded that poorer fact (detail) memory and to a greater extent decreased working memory capacity are the sources for age differences in constructing inferences.

Burke and Yee (1984) found that on a lexical decision task older adults still maintained the ability to access a word implied from a sentence (e.g., the cook cut the meat – implies the word “knife”), faster than one completely unrelated to the sentence. These results are in accordance with the conclusion that older adults are just as able to make inferences as their younger counterparts. Furthermore, the information required to draw these inferences was presented in a single sentence, eliminating the reliance on the retention of several facts. Light and Albertson (1988) added more fuel to the argument that older adults do not actually decline in their ability to reason. They compared younger and older age groups’ ability to make logical inferences, which demand an inference to be made (e.g., “Neil was forced to fly the plane” requires the inference that he flew the plane), versus pragmatic inferences, which do not force an inference to be made (e.g., “Bob was able to climb the ladder” does not mean that he did in fact climb the ladder). Pragmatic inferences are deemed to involve greater amounts of processing because they rely on accessing common knowledge about the way the world works, and if older adults have less processing resources they should perform more poorly on these types of inferences. However, the older adults did not differ in their ability to infer both
types of statements, thus rebuking the hypothesis that older adults' are unable to make
inferences due to diminished processing resources.

Findings from Reder, Wible, and Martin (1986) further supported the memory
difficulty hypothesis. Older and younger participants read short stories and answered
questions concerning them. Half of the participants were asked to judge whether the
plausible test statements were presented in the story (recognition task), thereby requiring
a direct retrieval strategy, or searching for specific facts. The other half of the
participants were asked to determine whether the test statements were plausible from the
information given in the story (plausibility task), relying on a plausibility strategy of
using available information to infer if a statement was true. In addition, half of the
plausible statements in each condition were explicitly stated in the story and half were
implicitly stated (the plausibility task included half plausible and half implausible
statements, while the recognition task included only plausible items). When the
plausibility of the statements was explicitly stated there were no age differences for both
the recognition and plausibility tasks. On the other hand, when the plausibility of the
statements had to be inferred, the older group did not differ from the younger group in
accurately responding whether the statement was plausible, but did in fact show poorer
performance in recognizing whether that statement had been shown in the story. It
appeared that for explicitly stated facts, using a plausibility strategy in the recognition
task resulted in accurate judgements, but applying this same strategy in the recognition
task to implied plausible statements resulted in erroneously responding that these
statements were seen before. Reder et al. concluded that older adults have difficulty
when an inference task requires retrieval of specific information, but not when the task is testing their ability to reason.

In sum, when task conditions place great demands on memory, older adults are unable to make correct inferences due to their inability to recall the relevant information (Light & Albertson, 1988). It appears that older adults have intact comprehension ability, but their decline in memory ability and resources are causing what at first glance appear to be differences in reasoning ability.

The conclusion that there are no age differences in the ability to actually form the inferences is further supported by Hamm and Hasher (1992). They investigated age differences in accepting or rejecting inferences that were central in understanding a passage. Younger and older adults read a series of short passages in which the final outcome was the same, but either expected (the first half of the passage was consistent with the final conclusion) or unexpected (the first half of the passage implied an interpretation that was different from the eventual outcome). When asked halfway through the expected passages, both age groups accepted the usual inference or outcome at the same rate, and were similarly unlikely to have the typical interpretation of the story when the outcome was unexpected. Therefore, older adults are not deficient in their ability to form inferences from stories. Unexpectedly though, Hamm and Hasher found older adults may entertain a broader array of inferences than younger adults, because they were far more likely to support both the typical and atypical inferences in the expected passages at the same time. Moreover, at the end of the unexpected stories, the older adults accepted the atypical interpretation more often than the younger adults, even though the outcome of the story was already demonstrated to be of a typical nature. It
appeared that the older adults had a tendency to hold both interpretations of the story in mind, even though the two interpretations were not simultaneously compatible with one another.

Hamm and Hasher proposed that older adults consider a broader range of possible interpretations when reading a passage, but this reasoning is at odds with the proposed diminished working memory hypothesis. If older adults have difficulty recalling inferences due to high memory demands, then they would be expected to produce fewer possible inferences, not more, than younger adults. Hamm and Hasher suggested that older adults’ enhanced entertaining of inferences may be due to faulty inhibitory mechanisms demonstrated by their difficulty in narrowing the range of possible interpretations. However, it could also be proposed that due to older adults’ limited memory of the story’s exact statements, they are simply considering more inferences and outcomes to the story which are plausible with the story’s main idea.

Although Hamm and Hasher (1992) concluded a different underlying mechanism for inferring difficulties with age, their research does help to confirm that the inferential changes that occur in aging are not caused by reasoning or language comprehension deficiencies. Therefore, older adults can draw inferences just as well as younger adults, and under simple task conditions, no age differences are found. However, when age differences do appear, the majority of inference and aging research supports the theory that these inferential deficits are really memory failures in disguise (Burke & Yee, 1984; Light & Albertson, 1988; Light et al., 1982; Reder et al., 1986). Age differences in making inferences typically only appear if working memory is taxed, and even then the
apparently worse inferring ability in older adults is a problem of recall of the relevant information, rather than one of actual difficulty in reasoning ability.

Therefore, the research literature supports the conclusions that the ability to draw inferences is retained with age. The ability to recall the main ideas of a story is similarly stable, but the recall of details is somewhat less accurate. As a result, older adults may be unable to discern what statements were actually included in a story (because they cannot remember the details) and those which were inferences of a story (because they cannot distinguish the inferred statements from the details). Consequently, older adults might rely on a plausibility strategy when responding to a story recognition task; if a statement is plausible from what they can recall about the gist of a story, older adults would respond that it was in fact a part of the story. Although by definition inferred statements are logical and consistent with the main ideas of a story, they have not actually been presented in the story. But due to their strategy to respond based on the plausibility of the statement with the gist of the story, older adults would be more likely to respond that an inference was in fact part of the original story. On the other hand, because of their superior memory ability, younger adults would still be able to use a direct retrieval strategy of the story items to differentiate among the inferences and the presented items (Reder et al., 1986).

Memory Differences within the Aged

The above hypothesis is consistent with the results comparing younger and older adults' ability to recall inferences (e.g., Hamm & Hasher, 1992; Light et al., 1982; Reder et al., 1986). However, due to the fact that recalling inferences appears to rest heavily on episodic memory ability, any further changes in this faculty could reasonably be expected
to also impact memory for inferences. In fact, episodic memory ability across the later part of the lifespan lacks the level of stability and uniformity typical of young adulthood. Evidence of this is apparent from three main points. First, cross-sectional data has found age differences within the older age range. For example, Nilsson (2003) found statistically significant differences in episodic memory between participants aged 55-60 and 65-70 years, and participants from each successive five-year cohort (e.g., 75-80) continued to significantly differ from the previous cohort. Similarly, Christensen, Mackinnon, Jorm, Henderson, Scott, and Korten (1994) found significant differences between participants aged 70-74, 75-79, and 80+ years on measures of crystallized and fluid intelligence and memory.

Second, there is longitudinal evidence that the rate of decline in episodic memory is greater in the latter half of old age compared to the first half. For example, over a 6-year period old-old individuals (75-84 and 85+ years) demonstrated more decline on an immediate word recall task than young-old individuals (65-74 years; Colsher & Wallace, 1991). Another longitudinal study by Small et al. (1999) tested young-old adults (55-70 years) and old-old adults (71-86 years) three times over 6 years on a story recall task. Although the young-old adults demonstrated a significant increase in performance over time, the old-old adults demonstrated a slight decline in story recall performance.

Finally, the cognitive aging process varies substantially across individuals, in that it does not affect all older adults the exact same way, or at a strict time or rate. While there are similarities and patterns in age-related cognitive changes, each individual has their own trajectory of change. It appears that a variety of variables other than age (e.g., genetic, biological, educational, health, and lifestyle) impact the changes that will occur,
and the weight of these factors varies across cognitive domains (Christenson, Mackinnon, Korten, Jorm, Henderson, Jacomb et al., 1999). The individualized nature of these factors and their complex interrelationships has resulted in older adults demonstrating far more interindividual variability, or differences from one another, in their performance on cognitive tasks than younger adults (Morse, 1993). Christensen et al. (1994) found greater interindividual variation with age in a large sample of older adults on measures of fluid intelligence and memory, and Hultsch, Hertzog, Dixon, and Small (1998) found that older individuals became less alike over a six-year period on seven of nine cognitive variables. Greater variability among older adults persists even when those with dementia or extreme scores are excluded, demonstrating that increased interindividual variability in older adults is not simply the result of a small number of participants who have greatly declined (Christensen et al., 1999), but a product of the diversity of the aging process.

As a result of these pieces of evidence, it is inappropriate to assume that all adults falling into an “older” age range perform similarly on any cognitive task. Previous research investigating inferences has only compared younger adults to groups of older adults spanning a wide range of ages (e.g., 65-95 years [Cohen, 1979]; 58-82 years [Light & Albertson, 1988]; 50-81 years and 56-86 years [Light et al., 1982]; 65-80 years [Reder et al., 1986]), or only focused on the younger end of the aged continuum (e.g., 62-75 years [Hamm & Hasher, 1992]; 63-73 years [Light et al., 1982]; 64-75 years [Reder et al., 1986]). However, the above factors strongly suggest that further differentiation among the aged may exist, and consequently, the ability to recognize inferences may differ within a cohort of older adults.
Story Memory and Alzheimer’s Dementia

Although some cognitive changes accompanying old age are to be expected in so-called “normal” aging, other changes are of a more pathological nature, particularly those associated with dementia. Dementia is a syndrome marked by global declines in cognitive functioning, particularly memory, from a previously higher level of intellectual ability, with a sustained onset over a period of months or years (Molloy & Lubinski, 1995). The severity of the syndrome is such that it affects the individual’s capacity to maintain employment, carry out daily activities, and in some cases, to care for themselves. While various types of dementia exist, the most prevalent is that associated with Alzheimer’s Disease (AD), also termed dementia of the Alzheimer’s type (DAT).

The most pronounced change in cognitive ability in AD is severe memory deficits, including declines in episodic and semantic memory, and delayed and immediate recall. However, the nature of these impairments in the early stage of the disease can look very similar to those that come with normal aging. For example, early AD patients and healthy older adults exhibit some overlap in their pattern of language change, including word finding problems, reduced information content, and a tendency to digress (Ulatowska, Allard, Donnell, Bristow, Haynes, Flower et al., 1988).

Cardebat, Démonet, and Doyon (1993) further investigated whether the narrative deficits in AD are qualitatively similar to those in normal aging. Old-old adults (75-89 years) and AD patients were shown a series of pictures that depicted a story, and asked to narrate the story with the pictures in full view. The most striking difference was the presence of narrative paraphasias, or brief micro-narratives of very loosely related but irrelevant material by the AD group. The old-old adults never displayed these deviations.
Moreover, while the normal adults mentioned the main characters of the story, only 75% of the AD participants did the same, and the AD patients produced far more secondary or irrelevant elements (to the story structure) than the normal adults. Cardebat et al. concluded that the unusual focus on irrelevant material by the AD patients demonstrated that AD may result in difficulty in establishing the hierarchy of relevance of the various story elements.

Similar results in support of this hypothesis were found by Ska and Guénard (1993). AD patients recalled less essential information and secondary details, and produced more irrelevant information than normal older adults when recalling a well-known fairy tale from memory. Also, Chapman, Ulatowska, King, Johnson, and McIntire (1995) had normal old adults (47-78 years), old-old adults (+80 years), and early stage AD individuals create a story about 3 different contextually rich pictures. Both groups of normal older adults interpreted the picture correctly more often, supplied more supporting information, and were better at integrating the information in a narrative form than the early AD patients. The literature appears to support the notion that AD patients have difficulty in identifying the hierarchy between the main ideas and details of a story (which is essential to exhibiting the levels effect), and consequently may differ quite dramatically from healthy older adults on a story recall task. In fact, on a delayed story recall task, very mild and mild AD participants produced fewer gist responses than healthy older adults, who did not differ from younger adults (Johnson, Storandt, & Balota, 2003).
However, AD is not a sudden, rapid change from normal memory abilities to dementia, and subtle cognitive impairments can be present years before the clinical diagnosis of AD. This preclinical stage of the disease is labelled mild cognitive impairment (MCI; according to Petersen, Smith, Waring, Ivnik, Tangalos, & Kokmen, 1999), during which individuals have only mild memory difficulties causing them to perform more poorly than others their age, but maintain otherwise normal cognitive functioning. Although these deficits allow MCI individuals to be identified as psychologically abnormal for their age group, they cannot yet be classified as indicating dementia. Therefore, on a continuum of cognitive intactness, MCI individuals occupy the transitional space between normal aging and the various stages of dementia. It is important to note that there are a number of different terms and diagnostic criteria used to classify MCI. For example, according to Petersen, Doody, Kurz, Mohs, Morris, Rabins et al. (2001), potential diagnoses include amnestic MCI (only memory problems), and MCI single or multiple (mild impairment in a single or multiple cognitive domains, with or without memory impairment), but the transitional state’s diagnosis is not limited to only these terms. In fact, a plethora of similar terms and classification schemes exists, but unfortunately none of these labels are universally accepted or applied (Tuokko & Frerichs, 2000). Despite this, the very existence of an intermediate stage in abnormal cognitive aging offers the prospect that timely interventions to delay losses in memory functioning and the progression to dementia may be possible. Consequently, the ability to identify MCI individuals early is critical, but the question remains as to which cognitive tasks or domains may have the highest diagnostic sensitivity and specificity.
Linn, Wolf, Bachman, Knoefel, Cobb, Belanger, et al. (1995) examined longitudinal data spanning 13 years of biennial neuropsychological testing on 1045 older adults. They found that cognitive deficits could be detected an average of 7 years before clinical diagnosis of AD, and more importantly, that the verbal learning and memory measures were the most sensitive predictors of later diagnosis. In another study, declines on tasks of memory and executive function 1.5 years prior showed the highest predictability of later AD (Chen, Ratcliff, Belle, Cauley, DeKosky, & Ganguli, 2001).

Collie and Maruff (2000) reported that the majority of studies investigating preclinical AD report deficits of verbal episodic learning and memory, as demonstrated by tasks such as story recall and verbal list learning. Furthermore, Small, Mobly, Jonsson Laukka, Jones, and Bäckman (2003) and Bäckman, Small, and Fratiglioni (2001) described deficits in episodic memory as the most consistent and pronounced manifestation of preclinical AD. For example, Bäckman et al. (2001) found that performance on a free word recall and recognition test predicted AD 3 and 6 years prior to diagnosis, while a measure of short-term memory (forward or backward digit-span) did not. In addition, verbal episodic memory performance in MCI individuals has been found to be as impaired as that of mild AD patients, even though the MCI group did not significantly differ from healthy older adults on measures of other cognitive domains (e.g., executive function; Petersen et al., 1999). Therefore, verbal episodic memory may be among the first cognitive domains to decline in AD, thereby offering the best predictive power for abnormal development (Amáiz & Almkvist, 2003).

It also appears that the cognitive deficits seen in preclinical AD are qualitatively similar, but quantitatively less severe than those prevalent in AD. Therefore, preclinical
AD individuals usually perform on an intermediate level between normal controls and AD patients (Collie & Maruff, 2000). Chapman, Zientz, Weiner, Rosenberg, Frawley, and Burns (2002) compared the qualitative discourse abilities of healthy older adults, mild AD individuals, and MCI adults. After listening to a story, the participants were asked to produce a summary of the story, recall the main idea, and formulate a lesson that could be learned from the story. These tasks served as a measure of gist-level processing, whereas detail-level processing was evaluated by recall and recognition tasks based on a particular detail of the story (i.e., the main character’s careers throughout his life). The mild AD and MCI individuals performed significantly lower on all measures of gist-level processing than the normal older adults, but the two impaired groups did not significantly differ from one another. However, it is important to note that while the MCI adults did not significantly differ from the AD individuals in their recall of gist items, just over half of MCI individuals actually scored in the impaired range on the main idea task. For detail-level processing, the normal older adults had the highest scores, the MCI individuals achieved intermediate scores, and the mild AD group performed the worst. Although the mild AD individuals had the poorest performance on the tasks of detail-level processing, all MCI adults showed impaired detail recall that was in the AD range. These results suggest that detail-level deficits may precede AD clinical symptoms, because all MCI individuals were impaired at this level but the healthy older adults were not. It also appears that AD is accompanied by a similar loss in gist-level processing which MCI individuals do not yet demonstrate. The unique decline of both gist and detail recall in AD is further supported from the result that performance on both the main idea and detail measures was able to robustly differentiate mild AD individuals from
normal older adults. Consequently, Chapman et al. proposed that changes in gist recall performance might act as a unique tool in diagnosing early AD, while changes in detail recall performance may be able to differentiate between MCI and healthy older adults.

In summary, verbal episodic memory may be one of the most sensitive domains to changes in the preclinical phase of AD (Collie & Maruff, 2000). While AD is marked by deficits in both gist and detail level recall, there is evidence that MCI individuals maintain gist-level processing but are as impaired in memory for details (Chapman et al., 2002). Due to the fact that even old-old adults in their 80s and 90s maintain gist memory (Ulatowska, Chapman, Highley, & Prince, 1998), but normal older adults do not decline as much as MCI individuals for detail-level processing (Chapman et al., 2002), MCI individuals would be expected to be able to recall fewer details and therefore inferences on a story recognition task compared to normal older adults.

Intraindividual Variability

All of the previous studies investigating age differences in story recognition and inferences have focused only on the level (typically assessed by mean accuracy) of the participants’ performance (e.g., Reder et al., 1986). However, Nesselroade (1991) has argued that an individual’s development is manifest in two types of change. Intraindividual change is defined as a relatively slow and enduring process, and is the fundamental producer of learning and development. Longitudinal research which compares an individual’s performance at one point in time to another point years later is evaluating this type of long-term change. On the other hand, intraindividual variability corresponds to relatively rapid and short-term changes within an individual, such as shifts in mood or arousal, and fluctuations in physical or cognitive performance.
Intraindividual variability has typically been considered an indicator of random error in performance, reflecting unreliability of measurement or "noise". However, recent research investigating intraindividual variability has consistently found that these short-term fluctuations represent systematic and lawful patterns of individual change (e.g., Hultsch, MacDonald, Hunter, Levy-Bencheton, & Strauss, 2000).

As a result, a particular focus of interest for intraindividual variability research became cognition, which is generally viewed as a relatively stable characteristic of a person. In fact, there is evidence of meaningful fluctuations in cognitive ability both moment to moment and week to week (e.g., Rabbitt, Osman, Moore, & Stollery, 2001), and measures of latency are particularly sensitive to this variability. Hultsch, MacDonald, and Dixon (2002) found that individuals who were more variable in their performance across various reaction time (RT) tasks were also more variable from trial to trial on an RT task, and individuals with increased intraindividual variability on one RT task were also more variable in their performance on other RT tasks. Hultsch et al. noted that correlations of variability across trials and across RT tasks are what one would expect to find if relatively stable internal mechanisms do underlie intraindividual variability rather than random error. This conclusion was validated further by the result that intraindividual variability in the nonverbal RT tasks (simple and choice reaction time) was a significant predictor of performance in other cognitive domains such as perceptual speed, episodic and working memory, and crystallized abilities. Hultsch and colleagues (2002) have labelled these fluctuations inconsistency, and it has been suggested by Li and Lindenberger (1999) and others that increased intraindividual variability may be an indicator of compromised neurological disturbance.
Greater Intraindividual Variability with Age

A critical implication for aging research is that these short-term changes appear more prevalent in older adults. The aged display a consistent pattern of increased intraindividual variability on cognitive tasks, and this continues to be true even if the actual level of performance among older and younger adults is similar. For example, Spieler, Balota, and Faust (1996) found that although the magnitude of the Stroop interference effect was nearly identical for younger and older adults, older adults were more inconsistent. In fact, any differences in mean level of performance for older and younger adults are independent of the variance accounted for by differences in inconsistency on cognitive tasks. Hultsch and his colleagues (2002) compared the RT performance of younger and older adults on four different cognitive tasks. Even after group differences in speed and practice were removed, the older adults were still significantly more inconsistent than the younger adults on all tasks.

Intraindividual variability in older adults’ performance has been demonstrated both across trials within a session (Spieler et al., 1996), and across multiple testing occasions. Hertzog, Dixon and Hultsch (1992) examined inconsistency by testing 7 older women weekly for 2 years on their ability to recall a story and found considerable intraindividual variability across these occasions. Similarly, Rabbitt et al. (2001) investigated older adults’ performance on six different tasks over 36 weekly testing sessions. Greater inconsistency across the trials was correlated with greater inconsistency across longer intervals such as days and weeks, demonstrating that intraindividual variability is a stable characteristic of individuals.
Not only do older adults display more intraindividual variability, its degree also appears to increase with age. Hultsch et al. (2002) investigated age differences on a battery of cognitive tasks, which included RT measures such as simple and choice reaction time (SRT and CRT), lexical decision, and semantic decision. They compared the performance of a group of younger adults (17-36 years) and three different older age groups, including young-old (54-64 years), mid-old (65-74 years), and old-old (75-94 years) adults. The old-old group was significantly more inconsistent than the other age groups for all four RT tasks, while significantly increasing inconsistency across the other age groupings varied depending on the RT task. Similar results were found by Williams, Hultsch, Strauss, Hunter, & Tannock (in press), who examined intraindividual variability in RT across the lifespan. They analyzed RT data from 273 participants, who were separated into seven different age groups ranging from early childhood (6-8 years) to elderly (60-81 years). Williams and his colleagues found inconsistency followed a U-shaped curve across the age distribution where children and older adults demonstrated higher levels of inconsistency than younger adults (18-29 years). It appears that throughout childhood, age is associated with decreases in inconsistency, reaching the lowest point during young adulthood, and thereafter increases with age to demonstrate higher levels of inconsistency.

This hypothesis has also been confirmed in the older part of the lifespan with longitudinal data spanning 6 years. MacDonald, Hultsch, and Dixon (2003) investigated whether intraindividual variability predicts changes in level of cognitive performance. The older participants were divided into the same three older age ranges, and completed the same 4 RT tasks as described above (Hultsch et al., 2002) in addition to cognitive
tests targeting perceptual speed, working memory, fluid reasoning, episodic memory, and crystallized verbal ability. Due to the fact that this sample was tested for 6 years, they represented a more select healthy group of older adults than the larger sample analyzed by Hultsch et al. (2002), which included individuals who later discontinued participation. However, MacDonald and colleagues were still able to replicate the Hultsch et al. (2002) findings that old-old adults are more inconsistent on RT tasks than young-old adults, demonstrating that increased inconsistency with age is not only a function of health. Moreover, the individuals who did not return for testing demonstrated more intraindividual variability than those who returned for all three sessions over 6 years, possibly predicting the participants’ eventual reason for dropping out (e.g., declining health, abnormal aging, or impending death). Next, MacDonald et al. found that intraindividual variability at the initial test occasions significantly predicted cognitive change for all measures, including story recall. Interestingly, the old-old adults showed significant increases in inconsistency across the 6-year period, while inconsistency remained stable or decreased slightly for the young-old and mid-old groups. Although the individual sessions of testing indicated higher amounts of inconsistency on the non-verbal RT tasks (SRT and CRT), as found in Hultsch et al. (2002), longitudinal increases were greater for the verbal RT tasks (lexical and semantic decision) than for the non-verbal RT tasks. MacDonald et al. proposed these differential increases in inconsistency may be due to the relatively later decline of crystallized abilities, which would have finally begun only in the old-old group. In addition, increasing longitudinal inconsistency was associated with declining cognitive performance for 5 of the 6 measures (all but vocabulary). Interestingly, inconsistency uniquely accounted for more
performance variability in the episodic memory measures than those involving basic processing domains. Finally, increased inconsistency predicted poorer cognitive performance uniformly across the entire adult age continuum, demonstrating that inconsistency is a relevant predictor of change throughout the entire aging process.

Previous research investigating story recognition and inferences has focused only on the level of performance (e.g., Reder et al., 1986), but recent studies have shown that intraindividual variability is a stable, systematic, and valuable indicator of an individual’s cognitive integrity (e.g., MacDonald et al., 2003). There is ample evidence from the inconsistency and aging literature that supports the idea that older adults demonstrate more intraindividual variability than younger adults, regardless of mean level of performance (e.g., Williams et al., in press). This age continuum also appears to continue throughout adulthood, with young-old adults significantly less inconsistent than the oldest of the old (e.g., Hultsch et al., 2002). Consequently, analyzing inconsistency on a story recognition task appears to be a worthwhile extension in story and inference research, and evidence strongly suggests that old-old adults will be more inconsistent on a story recognition task than young-old adults.

Intraindividual Variability and Older Clinical Populations

As noted earlier, intraindividual variability has consistently been shown to increase with age (Hultsch et al., 2002; MacDonald et al., 2003; Spieler et al., 1996; Williams et al., in press). However, one of the most intriguing possibilities that has emerged from the recent literature is the predictive ability of intraindividual variability to also differentiate groups with neurological disturbances.
Li and Lindenberger (1999) and others hypothesized that the presence of inconsistency represents compromised neurobiological mechanisms. If this is true, then performance by individuals with neurological disease or injury should demonstrate higher inconsistency than that of healthy individuals. In addition, if aging can also be attributed to compromised neurobiological mechanisms, then aging adults should be significantly more inconsistent than younger adults, yet still display less inconsistency in performance than abnormal adults. Finally, if inconsistency is evidence of impaired neurological factors and not more exogenous influences (e.g., fluctuations in stress or fatigue), then individuals who are neurologically intact but experience significant somatic disturbances (e.g., arthritis), should not display greater amounts of inconsistency than healthy adults.

Hultsch et al. (2000) investigated hypotheses of this sort in a comparison of three groups of older adults: healthy adults, adults with arthritis, and adults diagnosed with mild dementia. Their results were consistent with the above hypotheses: participants diagnosed with mild dementia demonstrated twice as much intraindividual variability as the neurologically intact participants, and adults with arthritis were not significantly more inconsistent than healthy adults. In addition, intraindividual variability also uniquely predicted neurological status independent of level of performance on the tasks. These results add evidence to the possibility that inconsistency does indeed result from neurological dysfunction rather than from more general health problems.

The diverse amounts of inconsistency were even found when this study format extended the measurement of inconsistency in normal and abnormal older adults from the cognitive domain to physical, sensory, and affective qualities (Strauss, MacDonald, Hunter, Moll, & Hultsch, 2002). Healthy older adults, older adults with a non-
neurological disturbance (arthritis), and older adults with diagnosed neurological compromise (dementia) were compared on their cognitive and physical performance, as well as their self-perceived affect and beliefs. First, it was hypothesized that if inconsistency on cognitive tasks is truly a result of underlying neurological dysfunction, it will be more highly correlated with inconsistency on physical tasks rather than on affect and belief measurements. Across the participant groups, intraindividual variability on the physical tasks uniquely predicted 53.5% of the latency and 82.6% of the accuracy variance in the cognitive tasks, while inconsistency in the affect/beliefs measures failed to make significant predictions. The lack of a significant prediction of cognitive performance by the self-perceived affect and belief measures' inconsistency adds evidence to the hypothesis that intraindividual variability is the result of neurological rather than exogenous factors. Next, inconsistency in physical function was positively correlated with the more demanding cognitive tasks (e.g., word and story recognition) for only those participants with dementia, while the non-demanding cognitive tasks (e.g., SRT) were correlated with physical performance inconsistency for all groups. The broader cognitive correlations for the demented adults with their inconsistent physical performance are consistent with the conclusion that inconsistency represents underlying neurological disturbance.

This hypothesis is even further validated by the findings that participants with different types of dementia vary in their amount of inconsistency. Older adults diagnosed with dementia with Lewy bodies demonstrated significantly more intraindividual variability on RT tasks within a trial, across trials, and across one week than AD or healthy participants (Walker, Ayre, Perry, Wesnes, McKeith, Tovee, et al., 2000).
Similarly, in comparison to older individuals with AD, those with frontal lobe dementia had significantly greater inconsistency in cognitive performance (Murtha, Cismaru, Waechter, & Chertkow, 2002).

Not only does the level of inconsistency fluctuate across the various types of dementia cases, but other disorders as well. Fuentes, Hunter, Strauss, and Hultsch (2001) found that persons with chronic fatigue syndrome showed higher levels of inconsistency than healthy individuals on each occasion of cognitive testing, even though these group differences diminished when inconsistency was compared across all occasions. In addition, consistent with the hypothesis that inconsistency is a result of overall brain deficits, older participants with AD and Parkinson’s disease were significantly more inconsistent within trials and across four occasions of cognitive measures than were healthy older adults (Burton, Strauss, Hultsch, Moll, & Hunter, 2004). However, the AD patients displayed more intraindividual variability than the Parkinson’s patients, suggesting that different diseases may impact consistency to varying degrees.

**Intraindividual Variability and Mild Cognitive Impairment**

If older adults with AD demonstrate more inconsistency than healthy older adults as an indication of their poorer neurological status (Hultsch et al., 2000), it remains possible that even minimal disturbances in the brain, such as those seen in MCI individuals, can be predicted by intraindividual variability. Dixon, Lentz, Garrett, MacDonald, Strauss, and Hultsch (2004) compared performance inconsistency in probable MCI individuals to that of healthy older adults, termed Neurologically Intact Controls (NIC). Probable MCI status was determined by being one or more standard deviations below their respective Age X Education group mean on at least one of five
cognitive measures (digit symbol, letter series, word recall, verbal fluency, and vocabulary). All other individuals were classified as NIC.

In their first experiment, the MCI group was more inconsistent than the NIC group on the RT tasks of lexical and semantic decision, but no significant differences were observed for either the simple or choice RT tasks. While the old-old groups of MCI and NIC adults did not differ in their intraindividual variability, the young-old and mid-old age groups of MCI were more inconsistent than their corresponding age healthy comparison groups. Also, the three age groups within MCI did not differ significantly from one another in their inconsistency, but the oldest NIC age group had higher levels of variability than the two other younger age groups. Therefore, the level of intraindividual variability in performance appears to be consistent across the MCI groups, regardless of age, and more interestingly that inconsistency level is similar to that displayed by healthy old-old adults.

The second experiment by Dixon and colleagues (2004) questioned whether the extent of impairment was related to intraindividual variability. In particular, they investigated whether persons with multiple domains of impairment (MCI-Moderate) were more variable in their performance than those with only one domain of impairment, and whether persons with only one domain of impairment (MCI-Mild) differed from NIC individuals in variability. A different sample of older adults was classified in a similar manner as the first group, but for this classification, if a participant was one standard deviation or more below their group mean level of performance on two or more cognitive measures, they were classified as MCI-Moderate. All participants were repeatedly tested on three RT tasks on 5 occasions, and for all three tasks, the MCI-Moderate group was
significantly more variable than the MCI-Mild group, who were more variable than the healthy older controls. In summary, those individuals with greater amounts of cognitive impairment (MCI-Moderate) demonstrated more intraindividual variability than those hypothesized to have less cognitive impairment (MCI-Mild).

This study demonstrates that higher levels of inconsistent performance are correlated with greater neurological disturbance, and that even subtle pathological changes in the brain can be reliably detected. More specifically, it showed that intraindividual variability may act as a potential early marker of preclinical dementia. The recent literature investigating inconsistency and clinical disorders offers support for the hypothesis that MCI individuals will demonstrate greater amounts of intraindividual variability on a story recognition task relative to older adults with No Cognitive Impairment (NCI).

**Task Differences in Intraindividual Variability**

Finally, the research literature is inconsistent as to whether the difficulty of a task influences the amount of intraindividual variability. West, Murphy, Armilio, Craik, and Stuss (2002) failed to find age differences in intraindividual variability on tasks which required minimal executive control (e.g., CRT), but older adults showed greater inconsistency on more demanding executive reaction time tasks (e.g., 1-back trials, which require an individual to recall and respond to the stimulus of the previous trial). West (2001) concluded that age-related increases in inconsistency are limited to those cognitive processes which are executive in nature, but this conclusion is not entirely supported by later research. Other studies have found just the opposite, showing significant age differences for tasks not requiring executive control (e.g., SRT and CRT,
Study 2, Dixon et al., 2004; SRT and CRT, Hultsch et al., 2002; CRT, Williams et al., in press), and these same tasks revealed greater age differences than more cognitively demanding tasks such as lexical or semantic decision (MacDonald et al., 2003). Despite these findings, the literature still remains inconsistent, as Study 1 by Dixon et al. (2004) failed to find significant age differences in SRT and CRT. Furthermore, the 1-back RT task demands substantially more of executive control than a lexical or semantic decision RT task, so comparisons between very non-executive tests (i.e., SRT and CRT) and only somewhat executive tests may not be the most conclusive. A more direct comparison found the greatest age effect on the 1-back RT task compared to the CRT and SRT tasks, and the cognitive status effect on the 1-back task was nearly four times that on SRT, and more than five times the effect on CRT (Study 2, Dixon et al., 2004). Therefore, it is apparent that the difficulty of a task may indeed play a role in the amount of inconsistency. Due to the fact that inferences are hypothesized to be more difficult to recognize, and one automatically draws them when comprehending a story, they may consequently exert a greater influence on the consistency of responding than the other statement types.

Present Study

To reiterate, research has shown aging results in both quantitative and qualitative declines in story recall. More specifically, older adults recall proportionately less details of a story than younger adults, but are relatively unimpaired in remembering the main ideas of the text. Another aspect of story recall, inference ability, remains relatively intact but may demonstrate declines if working memory is taxed and older adults are unable to recall the relevant information necessary to draw inferences. As a result, aging
and/or cognitive impairment may cause older adults to rely on a plausibility strategy in
story recognition, whereas younger and/or healthy adults are still able to directly retrieve
specific facts and items.

Research Questions

1. Are there Differences within Old Age in the Ability to Recognize Inferences?

The first issue of this study addresses age differences in story and inference
recognition within the stages of later life. Episodic memory appears to be a significant
factor in the ability to recognize inferences, but there is substantial evidence suggesting
that episodic memory and overall cognitive ability are not stable or uniform in the aged.
Research has found significant age differences between young-old and old-old adults on
episodic memory tasks (e.g., Nilsson, 2003), and there is longitudinal evidence that the
rate of decline in episodic memory is greater in the later half of old age (Small et al.,
1999). Moreover, the aging process has no definite path or speed, and appears to vary
according to a number of variables other than age, such as genetic or educational factors.
The individualized nature of the aging process has resulted in increased variability in
cognitive functioning, or fewer similarities when comparing one older individual to
another (e.g., Hultsch et al., 1998). Consequently, the assumption that all older adults
perform similarly on a cognitive task may be misleading. Thus far, research investigating
inferences has compared only younger and older adults (Burke & Yee, 1984; Cohen,
1979, 1981; Hamm & Hasher, 1992; Light & Albertson, 1988; Light et al., 1982), as has
research specifically investigating inferences and story recognition (Reder et al., 1986).
These pieces of evidence suggest that there may be further differentiation among older
adults in their ability to distinguish inferences on a story recognition task.
2. Does Cognitive Status Play a Role in the Ability to Recognize Inferences?

The second issue centers on the development of MCI and its possible differentiation from NCI older adults. One of the earliest changes that occurs in AD is the decline of verbal episodic memory. While the recall of both the main ideas and details of a story diminish in AD, during the preclinical phase of AD (MCI), recall for the most important points of a story is maintained but memory for details already exhibits decline (Chapman et al., 2002). However, old-old adults maintain gist memory (Ulatowska et al., 1998) and NCI older adults do not demonstrate as severe decline in memory for details as MCI individuals (Chapman et al., 2002). Consequently, the preservation of being able to recall the lower level propositions of a story and therefore inferences may uniquely differentiate MCI individuals from NCI older adults on a story recognition task.

3. Are there Differences within Old Age in the Amount of Inconsistency in Distinguishing Inferences?

The third question involves age differences in intraindividual variability for inferences. As mentioned earlier, research investigating story recognition and inferences has focused only on the participant’s level of performance or accuracy (e.g., Reder et al., 1986). However, recent studies have shown that there are stable and systematic patterns of one’s intraindividual variability on a cognitive task (e.g., Rabbitt et al., 2001), and inconsistency frequently acts as a more valuable predictive measure of cognitive ability than mean level of performance (e.g., Hultsch et al., 2002). There is ample evidence demonstrating that greater inconsistency accompanies increasing age (e.g., Williams et al., in press), and the rate of increase of inconsistency over time has been shown to be
greater in old-old adults (MacDonald et al., 2003). Finally, tasks which require greater executive control seem to also create less consistency in responding (e.g., Dixon et al., 2004), and inferences are more difficult to distinguish than other statement types. Consequently, inconsistency appears to be a valuable measure to investigate in story recognition, and there is good reason to suspect that old-old adults will be more inconsistent than young-old adults in story recognition, particularly for inferences.

4. Is Cognitive Status a Significant Factor in the Amount of Intraindividual Variability in Recognizing Inferences?

The final question addresses whether there are differences in the amount of inconsistency as a result of cognitive status. Research has shown that various groups with neurological disease have greater intraindividual variability than healthy older adults (e.g., AD, Hultsch et al., 2000; Parkinson’s disease, Burton et al., 2004), suggesting that inconsistency is indeed a marker of cognitive integrity. Greater intraindividual variability has also been found in individuals with subtle, preclinical cognitive deficits (i.e., MCI), and the greater the extent of preclinical impairment, the more inconsistent an individual is (Dixon et al., 2004). As mentioned earlier, the difficulty of a task appears to influence the degree of performance inconsistency (e.g., West et al., 2002), suggesting that inferences may elicit the greatest intraindividual variability of the statement types. These factors suggest that intraindividual variability on a story recognition task involving inferences will be able to differentiate MCI individuals from NCI participants, and even distinguish among severity of MCI status.
Method

This study used previously collected cross-sectional data obtained from an ongoing longitudinal study. The design of the study consists of multiple cross-sectional samples of older adults who are tested approximately biweekly on five occasions, and are also longitudinally retested at intervals of 1 year.

Participants

The participants were community-dwelling adults living in a medium-size metropolitan area (Victoria, British Columbia, Canada), recruited through advertisements to the general public specifically appealing for older adults who were concerned about their cognitive ability. Individuals with a self-reported history of extensive drug/alcohol abuse, major medical or neurological illnesses or injuries, psychiatric conditions requiring hospitalization, severe sensory impairment, or a Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) score less than 24 were excluded from the study. The final sample included 331 participants; however, due to various reasons including time and interest conflicts, 27 participants did not complete the first year of testing. Thus, data from a total of 304 participants (96 men and 208 women) ranging in age from 64 to 92, were analyzed in the present study.

Participants were categorized into two age groups spanning similar ranges to those described by previous research (e.g., MacDonald, Hultsch, & Dixon, 2003): A Young-Old (YO) age group (n = 170, 60 men, 110 women) aged 64 to 74 years (M = 69.67, SD = 2.74), and an Old-Old (OO) age group (n = 134, 36 men, 98 women), aged 75 to 92 years (M = 79.54, SD = 4.02).
In addition, the participants were classified into three groups reflecting their cognitive intactness. Cognitive status was determined by comparing each participant’s performance on five cognitive measures (Digit Symbol, Letter Series, Word Recall, Verbal Fluency, and Vocabulary) to the mean performance of an age and education stratified normative sample provided by the Victoria Longitudinal Study (VLS). Participants were classified as MCI if their performance was more than one standard deviation below their age and education-matched peers on one or more cognitive tasks. These individuals were furthered subdivided into MCI-Single (MCI-S) and MCI-Multiple (MCI-M) classifications. MCI-Single was defined as scoring more than one standard deviation below the normative sample on only one cognitive task, whereas MCI-Multiple meant the individual scored more than one standard deviation below the normative sample on two or more cognitive tasks. All other participants were classified as No Cognitive Impairment (NCI). These classifications were similar to those previously reported in Dixon et al. (2004). The NCI group included 136 participants (39 men, 97 women) ranging in age from 64 to 86, the MCI-Single consisted of 88 participants (23 men, 65 women) ranging in age from 65 to 88, and the MCI-Multiple group included 80 participants (34 men, 46 women) ranging in age from 66 to 92\(^1\) (see Table 1 for the \(n\) for each Age X Cognitive Status group).

Participants provided demographic and self-reported health information\(^2\) during an initial intake interview. In addition to the MMSE, several benchmark cognitive

---
\(^1\) One participant’s (aged 64) cognitive status was calculated with VLS norms based on a population aged 65-94.
\(^2\) Three measures of self-reported health were obtained: Self-rated health: Average of self-rated health compared to others of the same age, and compared to a perfect state of health on a 5-point scale (1= very poor to 5 = very good); Total health conditions: self-reported presence of 16 chronic health problems (e.g., diabetes, arthritis, high blood pressure); Total medications: number of prescription and non-prescription medications taken on a regular basis.
Table 1

Education, Health and Cognitive Variables as a Function of Age and Cognitive Status

<table>
<thead>
<tr>
<th>Variable</th>
<th>Young-Old</th>
<th></th>
<th>Old-Old</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCI (n=75)</td>
<td>MCI-S (n=54)</td>
<td>MCI-M (n=41)</td>
<td>NCI (n=61)</td>
</tr>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Years of education</td>
<td>16.24 (2.85)</td>
<td>15.43 (3.15)</td>
<td>15.12 (2.95)</td>
<td>14.79 (3.14)</td>
</tr>
<tr>
<td>Average health *</td>
<td>4.46 (0.54)</td>
<td>4.32 (0.50)</td>
<td>4.29 (0.62)</td>
<td>4.42 (0.56)</td>
</tr>
<tr>
<td>Total health conditions b</td>
<td>2.37 (1.75)</td>
<td>2.24 (1.50)</td>
<td>3.14 (1.92)</td>
<td>3.05 (1.76)</td>
</tr>
<tr>
<td>Total medications c</td>
<td>5.61 (3.71)</td>
<td>5.19 (2.93)</td>
<td>5.66 (3.34)</td>
<td>5.92 (3.27)</td>
</tr>
<tr>
<td>MMSE</td>
<td>29.28 (0.80)</td>
<td>28.94 (0.94)</td>
<td>28.73 (1.32)</td>
<td>28.72 (1.23)</td>
</tr>
<tr>
<td>WAIS Block Design (raw scores) d</td>
<td>41.64 (9.74)</td>
<td>36.94 (9.81)</td>
<td>31.49 (6.62)</td>
<td>35.74 (9.93)</td>
</tr>
<tr>
<td>WAIS Vocabulary (raw scores) e</td>
<td>59.41 (4.59)</td>
<td>56.59 (7.04)</td>
<td>53.85 (7.06)</td>
<td>58.05 (5.40)</td>
</tr>
<tr>
<td>WAIS estimated FSIQ</td>
<td>125.29 (10.66)</td>
<td>119.83 (11.06)</td>
<td>112.78 (10.65)</td>
<td>125.61 (11.64)</td>
</tr>
<tr>
<td>NAART errors f</td>
<td>10.65 (6.23)</td>
<td>15.74 (7.74)</td>
<td>16.12 (9.53)</td>
<td>12.67 (5.87)</td>
</tr>
<tr>
<td>Estimated NAART IQ</td>
<td>119.49 (4.86)</td>
<td>115.52 (6.03)</td>
<td>115.22 (7.43)</td>
<td>117.92 (4.58)</td>
</tr>
</tbody>
</table>

Note. NCI = No cognitive impairment; MCI-S = Mild cognitive impairment - Single; MCI-M = Mild cognitive impairment - Multiple.

MMSE = Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975); WAIS = Wechsler Adult Intelligence Scale-III (Psychological Corporation, 1997); FSIQ = full scale IQ; NAART = North American Adult Reading Test (Blair & Spreen, 1989).

*Average of self-rating of health compared to others of the same age, and compared to a perfect state of health (1 = very poor; 5 = very good)

*bTotal health conditions refers to self-reported presence of 16 chronic health problems (e.g., diabetes, arthritis, high blood pressure).

*cTotal medications represents number of prescription and non-prescription medications taken on a regular basis.

dOut of a possible total score of 68.

*eOut of a possible total score of 66.

*fOut of a possible total errors of 61.
measures were also obtained, including the Block Design and Vocabulary subtests of the Wechsler Adult Intelligence Scale-III (WAIS-III; Psychological Corporation, 1997), and the North American Adult Reading Test (NAART; Blair & Spreen, 1989). We computed estimates of full scale IQ (FSIQ) based on the age-adjusted Block Design and Vocabulary subtests (Sattler & Ryan, 1999) and premorbid IQ based on the NAART (Blair & Spreen, 1989).³

Group differences in each variable were evaluated by 2 (age) X 3 (cognitive status) X 2 (gender) ANOVAs. Table 1 shows the means for the demographic, health, and cognitive variables as a function of age and cognitive status.⁴ There were significant age group differences for education, $F(1, 292) = 11.32, p < .005, \eta^2 = .04$; total chronic health conditions, $F(1, 292) = 11.73, p < .005, \eta^2 = .04$; MMSE, $F(1, 292) = 21.97, p < .001, \eta^2 = .07$; WAIS Block Design, $F(1, 292) = 32.05, p < .001, \eta^2 = .10$; WAIS Vocabulary, $F(1, 292) = 17.23, p < .001, \eta^2 = .06$; and WAIS estimated FSIQ, $F(1, 292) = 4.54, p < .05, \eta^2 = .02$. The YO adults had significantly more education and reported fewer chronic conditions than the OO adults. In addition, the YO adults performed significantly better on the MMSE and WAIS Block and Vocabulary subsets, and had higher estimated FSIQs than the OO adults. There were no significant age difference on self-rated health, total number of medications, NAART errors, and NAART estimated IQ.

³ Blair and Spreen’s (1989) formula for estimating premorbid IQ using the NAART is $\text{NAART}_{\text{preIQ}} = 127.8 - 0.78 \times \text{NAART}_{\text{current}}$. This formula is based on the Wechsler Adult Intelligence Scale - Revised, whereas our estimate of current IQ is based on the WAIS-III, which is somewhat more difficult. Thus, the discrepancy between the NAART estimate and the WAIS-III estimate will, if anything, slightly overestimate cognitive decline.

⁴ Because there were few significant gender differences, and gender was not a main focus of the later analyses, means as a function of this variable were excluded from Table 1, and are reported in the text only where relevant.
There were significant differences among the cognitive status groups for all variables: education, $F(2, 292) = 5.68, p < .005, \eta^2 = .04$; self-rated health, $F(2, 292) = 3.19, p < .05, \eta^2 = .02$; total chronic health conditions, $F(2, 292) = 3.46, p < .05, \eta^2 = .02$; total medications, $F(2, 292) = 3.07, p < .05, \eta^2 = .02$; MMSE, $F(2, 292) = 10.59, p < .001, \eta^2 = .07$; WAIS Block Design, $F(2, 292) = 28.47, p < .001, \eta^2 = .16$; WAIS Vocabulary, $F(2, 292) = 29.31, p < .001, \eta^2 = .17$; WAIS estimated FSIQ, $F(2, 292) = 40.28, p < .001, \eta^2 = .22$; and NAART and NAART estimated IQ, $F(2, 292) = 22.04, p < .001, \eta^2 = .13$.

Post hoc tests using Tukey's pairwise comparisons ($p < .05$) revealed that the NCI group completed significantly more years of education than the MCI-M group and that the NCI adults rated themselves to be in better average health than did the MCI-M adults. The NCI and MCI-S groups did not differ significantly on either of these two measures. The MCI-M group also reported more chronic conditions than either of the two less cognitively impaired groups, who did not differ from one another. Despite the significant main effect for total medications, none of the contrasts revealed significant differences among the cognitive status groups. For the MMSE, the NCI and MCI-S adults obtained significantly higher scores than the MCI-M adults, but they did not differ from one another. Both of the WAIS subtests revealed a pattern of group differences that would be expected as the result of increasing cognitive impairment: The NCI adults performed better on the Vocabulary and Block Design subtests than the MCI-S and MCI-M adults, and the MCI-S adults performed better than the MCI-M adults. The same pattern of significant group differences was also seen for estimated FSIIQ, NAART errors, and the NAART estimated IQ.
In addition, there was a significant Age X Cognitive Status interaction on the NAART and NAART estimated IQ measures, $F(2, 292) = 6.87, p < .005, \eta^2 = .05$. To control for type I error, Bonferroni correction was applied to the resulting Age X Cognitive Status post-hoc analyses, setting $p < .01$. The analyses revealed that the YO/NCI adults had fewer errors on the NAART/higher NAART estimated IQ than the other two YO cognitive status groups, which did not differ. For the OO, the MCI-M adults had more errors on the NAART/lower NAART estimated IQ than both the NCI and MCI-S adults, who did not differ. In addition, the YO/MCI-M group performed better on the NAART/higher NAART estimated IQ than the OO/MCI-M group, whereas there were no age differences for the other two cognitive status groups.

Finally, there were significant gender effects for total chronic conditions, $F(1, 292) = 4.42, p < .05, \eta^2 = .02$, and total medications, $F(1, 292) = 6.26, p < .05, \eta^2 = .02$. Women ($M = 3.04$) reported more health problems than the men ($M = 2.67$) and women ($M = 6.10$) regularly took more medications than men ($M = 5.30$). In addition, there was a significant Age X Cognitive Status X Gender effect for education, $F(2, 292) = 3.09, p < .05, \eta^2 = .02$. Following Bonferroni’s correction which set $p = .004$, none of the remaining post hoc tests were significant, but the significant group interactions with education indicated that education should be used as a covariate in subsequent analyses.

**Measures**

The cognitive tasks used to classify MCI status covered multiple cognitive domains including perceptual speed, reasoning, episodic memory, verbal fluency, and vocabulary, and were administered only once in a group-administered test session. The
story recognition task was administered on a laptop computer and alternate versions of the task were administered over five individually-administered test sessions.

**Perceptual speed.** Perceptual processing speed was assessed using the WAIS-R Digit Symbol Substitution task (Wechsler, 1981). Participants were presented with a coding key pairing nine numbers (1 through 9) with nine symbols, below which were rows of randomly-ordered numbers with empty boxes. Participants were given 90 seconds to complete as many digit-symbol associations as possible by transcribing the appropriate corresponding symbols into the empty boxes below the given numbers. The number of correctly completed items represented the outcome measure.

**Reasoning.** Reasoning ability was assessed using the Letter Series test (Thurstone, 1962). Participants were presented with sets of strings of letters which followed a particular pattern. The task required deciphering the pattern in the target string and providing the next letter in the string congruent with the pattern presented. Participants were given 6 minutes to complete twenty strings of letters, and the outcome measure used was the total number of given letters correct.

**Episodic memory.** The episodic memory construct was measured by a word recall task which consisted of immediate free recall of 30 English words (Hultsch, Hertzog, & Dixon, 1990). The list consisted of 6 words from 5 taxonomic categories (e.g., birds, flowers) typed on a single page in unblocked order. Participants were given 2 minutes to study the list and 5 minutes to write their recall. The number of correctly recalled words was used as the outcome measure.

**Verbal fluency.** Participants’ verbal fluency was assessed using the Controlled Associations test (Ekstrom, French, Harman, & Dermen, 1976). The test required the
generation of as many synonyms as possible to a set of four target words. Participants were given 6 minutes to complete the test with the total number of correct synonyms representing the fluency score.

**Vocabulary.** Vocabulary was measured by a 36-item multiple-choice test (Ekstrom et al., 1976), where the object of the test was to select the correct definition of a target word from five possible definitions. Participants were given 10 minutes to complete the test with the total number of correct items representing the vocabulary score.

**Story Recognition.** Story recognition was based on five narrative stories selected from a set of 25 structurally equivalent texts developed by Dixon, Hultsch, and Hertzog (1989). Each story was approximately 300 words (24 sentences), 160 propositions long, and described events in the life (lives) of older adults (e.g., purchasing a new car, going camping). The stories were recorded on audiotape by a male professional actor and immediately following their single presentation, the recognition test was presented. The test consisted of 24 statements about the story shown one at a time on the computer screen. The 24 statements included 8 statements from each of the following three categories: True paraphrases - depicted ideas presented in the story; False paraphrases - depicted ideas not contained in the story; or Inferences - conveyed ideas consistent with the story, but not actually mentioned in the story (see Appendix for an example of a story and corresponding recognition tests statements). Participants were asked to press one of two keys to indicate whether the ideas presented in each statement were actually contained in the story (yes or no), and were reminded that exact wording of the story was

---

5 Any American locations contained in the stories were replaced with Canadian locations.
not important. The participants listened to a short example story and examples of the various types of statements that would appear before listening to the target story, and were reminded that because the ideas contained in the inference statements were not actually mentioned in the story, they should be responded to as though they were false. Participants were not explicitly instructed to focus on either accuracy or speed while completing the task. The outcome measures included both latency for each trial and the number of items correctly responded to. Only one story was presented per testing session.

Procedure

Participants completed multiple questionnaires, tests, and tasks focused on cognitive (e.g., memory) and non-cognitive (e.g., physiological) functioning. The various measures were administered across a group-administered test session, an individually-administered intake test session, and five individually-administered test sessions scheduled within a period of approximately 10 to 12 weeks. The story recognition task was one of many tasks administered in the individual sessions. The tasks across the five testing sessions were identical, and the order of the tasks was invariant across participants. Every attempt was made to ensure the sessions were between 10 to 18 days apart, however due to scheduling conflicts and holidays, more than 10 weeks was often required to complete the sessions. In order to minimize potential confounds, an effort was made to distribute the testing sessions across days of the week and the time of day (morning or afternoon). The group-administered test session took place at the University of Victoria, but a handful of participants unable to travel to the University were tested in their homes. For the majority of participants the intake test
session and the five testing sessions were administered in the participant’s own home, although a few individuals were tested at the University of Victoria.

Data Preparation

The participants were divided into one of six groups, determined by the mix of age (Young-Old/Old-Old) and cognitive status (NCI, MCI-Single, MCI-Multiple). Inspection of the data revealed some missing data; five participants did not complete one story and one participant failed to complete two of the stories because of equipment malfunction or tester error. To conserve as much data as possible, we imputed the missing accuracy and latency values for these participants by substituting the mean values for the appropriate age by cognitive status group on that story. This represented a conservative imputation which maintained group and interindividual differences in performance.

Although the five stories used in the study were structurally equivalent in terms of their average number of words, propositions, and reading/grade level required for their comprehension, they differed in their content. For example, the story describing the decision to purchase a new car involves more technical factors than the story describing the decision to move in with one’s adult child’s family. In addition, given that each participant has had different life experiences, the familiarity of the events depicted in the stories would also vary across participants. As a result of these factors, it is likely that there will be variation in performance across the five stories. However, because the stories were presented to participants in a fixed order, it is not possible to separate the story effects from practice or other order-related effects. Therefore, in order to maximize
external validity and the ability to extend this study’s results across story content, all analyses were collapsed across the five stories.

Results

The results are presented in five main parts. The first section focuses on the overall accuracy of story recognition performance. The second section reports the results of signal detection analyses that provide separate indicators of sensitivity (discriminability among stimuli) and bias in responding. In the third section we describe age and cognitive status group differences in latency of performance on the story recognition task. The fourth section examines intraindividual variability in performance to each statement type across the five occasions. Finally, correlational analyses are presented that examine the relationships between the accuracy, mean latency, and intraindividual variability variables.

As described earlier, there were significant group differences in educational attainment, and education was significantly correlated with overall accuracy ($r = .32, p < .001$) and approached significance for mean latency ($r = -.11, p = .06$). Thus, it was desirable to use the total years of education completed as a covariate in all of the following analyses. However, because it would be inappropriate to use a covariate if it interacted significantly with either of the between-subjects factors, we explored the relationship of education and these variables to overall accuracy and mean latency performance. Education did not significantly interact with age or cognitive status for either dependent measure, and therefore was used as a covariate in the subsequent analyses.
In the case of analyses where we had specific hypotheses, we report only the planned contrasts relevant to the hypotheses rather than the overall F-test results. In cases where there were no a priori hypotheses, the overall tests will be reported, followed by post hoc analyses.

**Overall Accuracy**

The overall accuracy measure referred to the number of correct responses given to each statement type out of a possible 40 correct responses. A Repeated Measures Analysis of Covariance (RM-ANCOVA) was performed with statement type as the repeated measures variable, and age and cognitive status as the between-subjects factors. The mean accuracy scores for the three statement types by age and cognitive status group are presented in Table 2.

As expected, there was a significant main effect of age, $F(1, 297) = 38.76, p < .001, \eta^2 = .12$; the YO ($M = 34.41$) were more accurate across all statement types than the OO ($M = 32.01$). With respect to statement type, we hypothesized that the participants would be less accurate to the inference statements than both the true and false paraphrases, which would not differ in accuracy. This hypothesis was in the expected direction; overall accuracy for the inference statements ($M = 28.03$) was lower than for the true [$M = 34.93; F(1, 297) = 34.59, p < .001, \eta^2 = .10$] and false statements [$M = 36.67; F(1, 297) = 51.14, p < .001, \eta^2 = .15$], and accuracy to the true and false statements did not differ.

We also hypothesized that age and statement type would interact. Specifically, the OO adults were expected to perform worse than the YO adults on the inference statements, but we did not expect any age group differences on the relatively easier true
Table 2

Mean Accuracy Scores to Each Statement Type by Age and Cognitive Status

<table>
<thead>
<tr>
<th>Statement Type</th>
<th>Young-Old</th>
<th></th>
<th></th>
<th>Old-Old</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCI</td>
<td>MCI-S</td>
<td>MCI-M</td>
<td>NCI</td>
<td>MCI-S</td>
<td>MCI-M</td>
</tr>
<tr>
<td>True Paraphrase</td>
<td>M</td>
<td>35.85</td>
<td>35.09</td>
<td>35.27</td>
<td>35.55</td>
<td>34.54</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(2.59)</td>
<td>(2.80)</td>
<td>(2.83)</td>
<td>(2.50)</td>
<td>(3.56)</td>
</tr>
<tr>
<td>False Paraphrase</td>
<td>M</td>
<td>38.52</td>
<td>37.75</td>
<td>36.88</td>
<td>37.5</td>
<td>36.69</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(1.48)</td>
<td>(2.37)</td>
<td>(2.64)</td>
<td>(2.19)</td>
<td>(3.07)</td>
</tr>
<tr>
<td>Inference</td>
<td>M</td>
<td>32.52</td>
<td>30.46</td>
<td>27.37</td>
<td>28.16</td>
<td>26.95</td>
</tr>
<tr>
<td></td>
<td>(SD)</td>
<td>(5.28)</td>
<td>(7.24)</td>
<td>(7.20)</td>
<td>(7.16)</td>
<td>(9.09)</td>
</tr>
</tbody>
</table>

Note. Total possible score per statement type = 40. NCI = No cognitive impairment; MCI-S = Mild cognitive impairment - Single; MCI-M = Mild cognitive impairment - Multiple.
and false statements. Figure 1 illustrates this interaction which was generally supported by the results. Both age groups demonstrated poorer performance on the inference statements relative to the true and false statements, the OO showed a significantly greater drop in performance on the inference statements relative to the true and false statements ($M = 8.99$) than the YO ($M = 6.05$; $M_{\text{diff}} = 2.93$, $SEM_{\text{diff}} = 0.82$, $p < .001$). However, the YO were more accurate than the OO for all three statement types [True: $F(1, 297) = 7.55$, $p < .01$, $\eta^2 = .03$; False: $F(1, 297) = 34.00$, $p < .001$, $\eta^2 = .10$; Inference: $F(1, 297) = 18.25$, $p < .001$, $\eta^2 = .06$]. Figure 1 also demonstrates an unexpected greater relative advantage for the YO in the ease of accurately responding to the false statements compared to the true statements. A posthoc test confirmed this, as the YO showed a larger relative advantage for the false statements ($M = 2.41$) than the OO ($M = 1.26$; $M_{\text{diff}} = 1.15$, $SEM_{\text{diff}} = 0.47$, $p < .05$).

We expected the effect of cognitive status on overall accuracy to reveal that increasing cognitive impairment would result in poorer overall accuracy. This was confirmed, as the NCI group ($M = 34.69$) was more accurate than the MCI-S ($M = 33.58$; $M_{\text{diff}} = 1.03$, $SEM_{\text{diff}} = 0.40$, $p < .05$) and MCI-M groups ($M = 31.37$; $M_{\text{diff}} = 3.06$, $SEM_{\text{diff}} = 0.41$, $p < .001$), and the MCI-S group was more accurate than the MCI-M group ($M_{\text{diff}} = 2.03$, $SEM_{\text{diff}} = 0.45$, $p < .001$). The Cognitive Status X Statement Type interaction was significant, $F(4, 592) = 2.73$, $p < .05$, $\eta^2 = .02$ (see Figure 2). Posthoc analyses using Bonferroni’s correction ($p < .005$) revealed the following significant differences among the groups for the three statement types: for true statements, the NCI were more accurate than the MCI-M ($M_{\text{diff}} = 1.42$, $SEM_{\text{diff}} = 0.42$, $p < .005$); for false statements, the NCI and MCI-S were more accurate than the MCI-M (NCI: $M_{\text{diff}} = 3.23$,}
Figure 1

Mean Overall Accuracy by Age and Statement Type
Figure 2

Mean Overall Accuracy by Cognitive Status and Statement Type
SEM_{diff} = 0.41, p < .001; MCI-S: M_{diff} = 2.50, SEM_{diff} = 0.45, p < .001); and for inferences, the NCI and MCI-S were more accurate than MCI-M (NCI: M_{diff} = 5.45, SEM_{diff} = 1.04, p < .001; MCI-S: M_{diff} = 3.99, SEM_{diff} = 1.14, p < .005). Finally, as can be seen in Figure 2, the magnitude of the relative drop from true/false accuracy to inference accuracy appeared to increase with increasing cognitive impairment, but the only significant difference was that the NCI group had a relatively smaller drop in accuracy performance (M = 6.33) than the MCI-M group (M = 9.45; M_{diff} = 3.13, SEM_{diff} = 1.01, p < .01). Similarly, it appeared that it was easier to accurately respond to the false rather than true statements, but the size of this relative advantage appeared to differ based on cognitive status. Posthoc tests (Bonferroni’s correction: p < .03) confirmed that the NCI and MCI-S adults did not differ from one another in the size of this relative advantage (NCI: M = 2.35; MCI-S: M = 2.46), and this advantage was significantly greater than for the MCI-M adults (NCI vs. MCI-M: M_{diff} = 1.81, SEM_{diff} = 0.57, p < .01; MCI-S vs. MCI-M: M_{diff} = 1.92, SEM_{diff} = 0.63, p < .01), for whom the advantage nearly disappeared (M = 0.54).

**Signal Detection Analyses**

Signal detection procedures were also performed on the accuracy data to provide a more comprehensive picture of story recognition performance. The overall accuracy analysis only examined the number of correct responses given for each statement type without taking into account individual differences in criterion responding. For example, if Participant A achieves a higher accuracy score to the true paraphrases than Participant B, one interpretation may be that Participant A recalled more information from the story. However, if it is also the case that Participant A performs more poorly than Participant B
on the false paraphrases, the initial interpretation no longer holds true. It appears that Participant A may have had a lower criterion level for responding “yes” to a statement, thus creating the illusion of superior ability. Consequently, examination of these two pieces of information, hits and false alarms, provides a clearer illustration of the participant’s discriminability and true accuracy on the task.

The typical signal-noise situation on which signal detection theory is based has only one signal curve and one noise curve. However, because the story recognition task had two potential noise curves (false paraphrases and inferences), two signal detection analyses were conducted for each participant. The first analysis used the number of correct responses to the true paraphrases (responding “yes”) as hits and the number of incorrect responses to the false paraphrases (responding “yes”) as false alarms. The second signal detection procedure again involved the number of hits to the true paraphrases, but included the number of incorrect responses to the inferences (responding “yes”) as false alarms. For each participant for each analysis, $d'$, representing sensitivity (how strongly an observer reacts to a signal), and criterion (minimum level of activation necessary in an ambiguous situation for an observer to claim detection of a signal) were calculated from the hit rate and the false alarm rate. Larger values of $d'$ indicate higher sensitivity to the difference between the signal and noise distributions. A $d'$ near zero indicates chance performance. In the case of the criterion measure, a value of zero indicates no bias, a positive value indicates conservativeness, and a negative value refers to liberalness or higher riskiness. For both analyses, the overall accuracy scores to each

---

6 It could be argued that the $\ln(\Delta)$ measure, which also provides a measure of bias in signal detection theory, could also have been used in this circumstance compared to the criterion measure. This possible argument was taken into consideration and $\ln(\Delta)$ was calculated for both analyses for each participant using the formula $\exp(d' \ast (2 \ast - z - false\ alarms - d') / 2)$, and then the $\ln$ of $\Delta$ was taken. The resulting
statement type were converted to proportions (hits for true statements) or 1 - proportions (false alarms for false and inference statements), and then changed to z-scores. For both measures of sensitivity, it was hypothesized that increasing cognitive impairment would result in decreased sensitivity to the difference between the statement types. Also, cognitive status and age were expected to interact, such that OO/MCI-M participants would perform far worse overall than any of the other groups. Group differences in criterion were not expected for either signal detection analysis.

**True/False**

$d'_{(True/False)}$

The $d'$ that corresponded to the sensitivity to the difference between the true and false statements was calculated using the formula: $z\text{-hits(True)} - z\text{-false alarms(False)}$.

An ANCOVA was performed with the between-subjects factors of age and cognitive status. As expected, there was a significant main effect of age, $F(1, 297) = 29.69, p < .001, \eta^2 = .09$; the YO adults ($M = 3.03$) were more sensitive to the difference between the true and false statements than the OO adults ($M = 2.63$). Planned contrasts for cognitive status were consistent with expectations; the NCI ($M = 3.08$) group was more sensitive to the difference between the true and false statements than the MCI-S group ($M = 2.87; M_{\text{diff}} = 0.21, \text{SE}M_{\text{diff}} = 0.08, p < .01$) and the MCI-M group ($M = 2.44; M_{\text{diff}} = 0.60, \text{SE}M_{\text{diff}} = 0.08, p < .001$), and the MCI-S group was more sensitive to the difference than the MCI-M group ($M_{\text{diff}} = 0.38, \text{SE}M_{\text{diff}} = 0.09, p < .001$). Figure 3 illustrates the significant Age X Cognitive Status interaction, and planned contrasts confirmed the expected outcome: For the YO, the NCI group was more sensitive to the difference

---

**ANCOVAs** revealed the exact same effects as the two criterion measures, and consequently only the criteria were reported.
Cognitive Status

Figure 3

Mean $d'$ (True/False) as a Function of Age and Cognitive Status
between the true and false statements than the MCI-M group ($M_{\text{diff}} = 0.32$, $\text{SEM}_{\text{diff}} = 0.11$, $p < .005$), whereas for the OO, both the NCI and the MCI-S groups had higher sensitivity than the MCI-M group (NCI: $M_{\text{diff}} = 0.88$, $\text{SEM}_{\text{diff}} = 0.12$, $p < .001$; MCI-S: $M_{\text{diff}} = 0.66$, $\text{SEM}_{\text{diff}} = 0.14$, $p < .001$). Also, the YO were more sensitive than the OO to the true/false difference, but only if they were classified as NCI [$F(1, 133) = 6.77$, $p < .05$, $\eta^2 = .05$] or MCI-M [$F(1, 77) = 22.77$, $p < .001$, $\eta^2 = .23$]. However, the effect size for the MCI-M age difference was nearly five times greater than for the NCI group, demonstrating that there was indeed a greater drop in sensitivity from the combination of older age and poorer cognitive status.

**Criterion (True/False)**

The criterion for the true/false analysis was computed using the equation: $\text{ABS}[z_{\text{false alarms(False)}}] - (d' (\text{True/False}) / 2)$. An age by cognitive status ANCOVA showed significant main effects for both factors, Age: $F(1, 297) = 3.96$, $p < .05$, $\eta^2 = .01$; Cognitive Status: $F(2, 297) = 5.55$, $p < .005$, $\eta^2 = .04$; but the interaction was not significant. The YO ($M = 0.22$) were more conservative in responding to the true/false distinction than the OO ($M = 0.12$). Post-hoc analyses of the cognitive status effect using Tukey’s pairwise comparisons and Bonferroni’s correction ($p < .02$) showed that the MCI-M group ($M = 0.06$) was less conservative than the two less more cognitively intact groups (NCI: $M = 0.22$; $M_{\text{diff}} = 0.16$, $\text{SEM}_{\text{diff}} = 0.04$, $p < .005$; MCI-S: $M = 0.21$; $M_{\text{diff}} = 0.15$, $\text{SEM}_{\text{diff}} = 0.05$, $p < .005$), which did not differ.
The second d', involving the true and inference statements, was calculated using the formula: $z_{hits(True)} - z_{false \ alarms(Inference)}$, and an Age X Cognitive Status ANCOVA was applied to the data. The main effects of age and cognitive status revealed the same pattern as the d' (True/False) measure: the YO ($M = 2.14$) were more sensitive to the difference between the two statement types than the OO ($M = 1.68$, $F(1, 297) = 30.00$, $p < .001$, $\hat{\eta}^2 = .09$); and planned contrasts for cognitive status showed the NCI group ($M = 2.15$) was more sensitive to the difference between the true and inference statements than the MCI-S ($M = 1.96$; $M_{diff} = 0.20$, $SEM_{diff} = 0.09$, $p < .05$) and the MCI-M groups ($M = 1.55$; $M_{diff} = 0.54$, $SEM_{diff} = 0.09$, $p < .001$), and the MCI-S group was more sensitive than the MCI-M group ($M_{diff} = 0.34$, $SEM_{diff} = 0.10$, $p < .001$). In addition, as can be seen in Figure 4, the pattern of results for the hypothesized Age X Cognitive Status interaction was also similar to the analysis of the first d' measure. For the YO, only the NCI group was more sensitive to the distinction than the MCI-M group ($M_{diff} = 0.43$, $SEM_{diff} = 0.12$, $p < .001$), but for the OO, both the NCI and MCI-S groups had greater sensitivity than the MCI-M group (NCI: $M_{diff} = 0.66$, $SEM_{diff} = 0.13$, $p < .001$; MCI-S: $M_{diff} = 0.48$, $SEM_{diff} = 0.15$, $p < .005$). Moreover, for all three cognitive status groups, the YO were more sensitive than the OO [NCI: $F(1, 133) = 14.29$, $p < .001$, $\hat{\eta}^2 = .10$; MCI-S: $F(1, 85) = 4.69$, $p < .05$, $\hat{\eta}^2 = .05$; MCI-M: $F(1, 77) = 13.91$, $p < .001$, $\hat{\eta}^2 = .15$], but as expected, the greatest effect size was for the MCI-M group, again showing the interactive detrimental effect of older age and increased cognitive impairment.
Mean $d'$ (True/Inference) as a Function of Age and Cognitive Status

Figure 4
Criterion (True/Inference)

The criterion measure for the second analysis was computed using the similar formula: \( \text{ABS}[z\text{-false alarms(Inference)}] - (d' (True/Inference) / 2) \). There were no significant main effects or interaction effects resulting from the Age X Cognitive Status ANCOVA.

Difference between \( d' \) (True/False) and \( d' \) (True/Inference)

Although the \( d' \) measures provide an illustration of possible group differences in sensitivity, a comparison between performance on each measure can further indicate the difficulty of the two statement comparisons relative to one another. It was hypothesized that it would be easier to differentiate between the true and false statements than the true and inference statements. Consequently, by comparing the outcome signal detection measures from the two analyses we can get a quantitative representation of the increase in difficulty from true/false to true/inference. Furthermore, it may be the case that the size of the increase in difficulty may vary as a function of age and cognitive status. It was expected that Old-Old participants and those with greater cognitive impairment would experience a larger increase in difficulty from the true/false distinction to the true/inference one. In addition, it was expected that like the individual \( d' \) analyses, there would be an interaction between age and cognitive status such that OO/MCI-M group would demonstrate a larger drop in sensitivity than all other groups.

An Age X Cognitive Status X \( d' \) RM-ANCOVA revealed that the two \( d' \)s significantly differed from one another, \( F(1, 297) = 46.78, p < .001, \eta^2 = .14 \). The \( d' \) (True/Inference) score (\( M = 1.87 \)) was smaller than the \( d' \) (True/False) score (\( M = 2.79 \)), demonstrating that it was indeed more difficult to differentiate between the true and
inference statements than the true and false paraphrases. For each participant, we made a comparison between the two d's by subtracting the $d'(\text{True/Inference})$ from the $d'(\text{True/False})$. Figure 5 shows the mean $d'(\text{True/False})$ and $d'(\text{True/Inference})$ by age group, and Figure 6 shows the comparison by cognitive status. These data were analyzed by an Age X Cognitive Status ANCOVA, but failed to reveal the expected main effects. In addition, planned contrasts for the hypothesized Age X Cognitive Status effect were not consistent with expectations (see Figure 7): The only significant difference among the various groups was that for the OO, the NCI group ($M = 1.02$) had a larger drop in sensitivity than the MCI-M group ($M = 0.79; M_{\text{diff}} = 0.22, \text{SEM}_{\text{diff}} = 0.10, p < .05$), perhaps demonstrating that the OO/MCI-M group already had such low sensitivity to the first type of comparisons (true/false) that they could not drop much further in sensitivity when they attempted to make the second more difficult comparisons (true/inference).

**Difference between criterion (True/False) and criterion (True/Inference)**

A more difficult comparison like true/inference may result in a shift towards greater liberalism in responding because an individual will likely have less certainty in their response compared to making a distinction between a true and a false paraphrase. Consequently, it was expected that there would be a significant difference between the two criterions, but there were no hypotheses concerning any group differences in the change from criterion (True/False) to criterion (True/Inference).

A RM-ANCOVA revealed that the difference in criterion for responding to the true/false ($M = 0.16$) and true/inference ($M = -0.30$) distinctions was significant, $F(1, 297) = 46.78, p < .001, \theta^2 = .14$, in that the more difficult true/inference discrimination
Figure 5

Mean $d'$ (True/False) and $d'$ (True/Inference) as a Function of Age
Figure 6

Mean $d'$ (True/False) and $d'$ (True/Inference) as a Function of Cognitive Status
Figure 7

Difference Between $d'$ (True/False) and $d'$ (True/Inference) by Age and Cognitive Status
caused participants to be more liberal in their likelihood of responding yes to an ambiguous statement. The group means for both criterion analyses are shown by age group in Figure 8 and by cognitive status in Figure 9. Similar to the calculation for the difference between the d's, for each participant the criterion (True/False) was subtracted from the criterion (True/Inference). The only significant effect from the Age X Cognitive Status ANCOVA was the interaction of Age X Cognitive Status, $F(1, 297), p < .05, \eta^2 = .02$. To control for Type I error, Bonferonni’s correction set the significance value at $p < .01$, but none of the resulting post-hoc tests were significant.

**Mean Latency**

For the mean latency measure, the RT (ms) to the 40 trials of each statement type were totalled and averaged for each participant. This measure reveals any prior group differences in response time, and is an essential precursor to demonstrating and understanding the statistical technique used to obtain the measures of intraindividual variability. Mean latency was analyzed using an RM-ANCOVA with statement type as the repeated measures variable, and age and cognitive status as the between-subjects factors. The mean latency scores to each statement type as a function of age and cognitive status are shown in Table 3.

The expected main effect of age for mean latency was significant, $F(1, 297) = 25.29, p < .001, \eta^2 = .08$; the OO adults ($M = 5437.26$) were slower to respond than the YO adults ($M = 4709.28$). For cognitive status, it was expected that the MCI-M group would be slower than the MCI-S group, which, in turn, would be slower than the NCI group. Planned contrasts revealed the MCI-M adults ($M = 5737.27$) had higher mean latencies than the NCI ($M = 4620.32$, $M_{diff} = 1119.63$, $SEM_{diff} = 171.40, p < .001$) and
Age Group

Figure 8

Mean criterion (True/False) and criterion (True/Inference) by Age
Cognitive Status

Note. The criterion (True/False) for the MCI-Multiple group is not significantly different from zero.

Figure 9

Mean criterion (True/False) and criterion (True/Inference) by Cognitive Status
Table 3

Mean Latency Scores to Each Statement Type by Age and Cognitive Status

<table>
<thead>
<tr>
<th>Statement Type</th>
<th>Young-Old</th>
<th>Old-Old</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCI</td>
<td>MCI-S</td>
</tr>
<tr>
<td>True Paraphrase</td>
<td>M</td>
<td>3968.25</td>
</tr>
<tr>
<td>(SD)</td>
<td>(979.37)</td>
<td>(1194.82)</td>
</tr>
<tr>
<td>False Paraphrase</td>
<td>M</td>
<td>3779.53</td>
</tr>
<tr>
<td>(SD)</td>
<td>(863.52)</td>
<td>(804.63)</td>
</tr>
<tr>
<td>Inference</td>
<td>M</td>
<td>5359.98</td>
</tr>
<tr>
<td>(SD)</td>
<td>(2018.36)</td>
<td>(2201.24)</td>
</tr>
</tbody>
</table>

Note. Times are in milliseconds. Mean latency was calculated across 40 trials of each statement type. NCI = No cognitive impairment; MCI-S = Mild cognitive impairment - Single; MCI-M = Mild cognitive impairment - Multiple.
MCI-S adults ($M = 4862.21; M_{diff} = 876.97, SEM_{diff} = 188.55, p < .001$), but the latter two groups did not differ. Age and cognitive status were expected to significantly interact whereby the OO/MCI-M group would have the largest mean latency. Planned contrasts revealed a pattern consistent with this hypothesis as the following contrasts were significant: For the YO, the NCI group ($M = 4369.25$) was faster in responding than the MCI-M group ($M = 5071.14; M_{diff} = 717.54, SEM_{diff} = 227.44, p < .005$), but for the OO, both the NCI ($M = 4871.39$) and MCI-S groups ($M = 5036.97$) were faster than the MCI-M group ($M = 6403.41$; NCI: $M_{diff} = 1518.00, SEM_{diff} = 260.30, p < .001$; MCI-S: $M_{diff} = 1351.05, SEM_{diff} = 297.55, p < .001$). Also, the YO were faster than the OO only for the NCI [$F(1, 133) = 5.83, p < .05, \eta^2 = .04$] and MCI-M cognitive status groups [$F(1, 77) = 18.36, p < .001, \eta^2 = .19$], but the effect size for the age difference for the MCI-M group was nearly five times that for the NCI group.

With respect to statement type, we hypothesized that the greater relative difficulty of the inferences would cause the participants to be slower at responding to them compared to the true and false paraphrases, which would not differ in response time. However, no significant differences were found. The Age X Statement Type interaction was significant, $F(2, 296) = 6.24, p < .005, \eta^2 = .04$ (see Figure 10), but posthoc tests (Bonferroni’s correction: $p < .01$) revealed that for within each age group there were no significant differences in the mean RT to the true, false, and inferred statements. The YO were faster than the OO for each statement type, [true: $F(1, 297) = 16.91, p < .001, \eta^2 = .05$, false: $F(1, 297) = 45.78, p < .001, \eta^2 = .13$, inference: $F(1, 297) = 6.89, p < .01, \eta^2 = .02$], but we must look to the effect sizes of these posthoc tests to explain the significant interaction. The greater age difference in responding to the false statements illustrates
Figure 10

Mean Latency as a Function of Statement Type and Age
that although the false statements are the quickest and therefore easiest to respond to for
the YO, this advantage disappears for the OO.

A similar pattern was found for the significant Cognitive Status X Statement Type
interaction, $F(4, 592) = 2.51, p < .05, \eta^2 = .02$, as shown in Figure 11. Posthoc analyses
(Bonferroni's correction: $p < .005$) within each group failed to find any significant
differences among the statement types. However, the contrasts across the cognitive status
groups were significant: for the true statements, the NCI and the MCI-S groups were
faster than the MCI-M group (NCI: $M_{diff} = 916.09, SEM_{diff} = 153.23, p < .001$; MCI-S:
$M_{diff} = 540.96, SEM_{diff} = 168.56, p < .005$); for the false statements, the NCI and MCI-S
groups were faster than the MCI-M group (NCI: $M_{diff} = 1213.88, SEM_{diff} = 165.94, p <
.001$; MCI-S: $M_{diff} = 989.34, SEM_{diff} = 182.54, p < .001$); and for the inferences, the NCI
and MCI-S groups were faster than the MCI-M group (NCI: $M_{diff} = 1228.92, SEM_{diff} =
317.27, p < .001$; MCI-S: $M_{diff} = 1100.59, SEM_{diff} = 349.01, p < .005$). Again, the size of
the mean differences are left to explain this interaction. Although the pattern of
significant cognitive status differences was the same for each statement type, the mean
differences between the MCI-M group and the NCI and MCI-S groups grows
successively larger from the easier true statements to the more difficult inference
statements, with the largest increase from the true to false statements. In addition, despite
no significant differences among the statement types for any of the cognitive status
groups, Figure 11 does seem to suggest that relative to their own responding, the MCI-S
group had a small advantage in responding to the false statements, the MCI-M group
actually took longer to respond to them relative to the true statements.
Cognitive Status

Figure 11

Mean Latency as a Function of Statement Type and Cognitive Status
Intraindividual Variability

Data Purification

As noted previously, intraindividual variability refers to the within-person changes in responding that are independent of systematic time-related effects (e.g., practice). Following the procedures described by Hultsch et al. (2000), several steps must be completed in order to obtain this “pure” measure of variability. The distributions of raw latency scores were first examined for outliers. Extremely fast or slow responses most likely represent various types of errors (e.g., accidental key press, distraction of participant), and were subsequently eliminated from the proposed analyses thereby reducing error variance. The outliers were determined by establishing upper and lower bounds and trimming any scores outside of these boundaries. The lower limit on the task was based on previous research by Hultsch et al. (2000), who used 1,000 ms as the lower boundary for legitimate responses on the story recognition task. This limit was applied to each response, regardless of story, statement type or Age X Cognitive Status group. However, as previously demonstrated, there were significant group differences in mean latency of responding as a function of age, cognitive status, and statement type. Therefore, the upper boundaries were calculated by computing the mean and standard deviation (SD) for each of the six Age X Cognitive Status groups by three statement types (18 upper limits) and accepting only those responses within +3 SDs of the mean. The number of responses that exceeded these upper and lower boundaries was relatively small: True = 1.9%, False = 1.0%, Inference = 0.5%.

7 Despite the potential varying familiarity among the participants with the topics of the five stories, there is no reason to expect that responses to one story will be faster than responses to another story, as this advantage or disadvantage would be reasonably expected to affect only the accuracy in responding.
In order to avoid statistical problems associated with missing data, missing value estimates were imputed for each statement type based on regression analyses using the relationship among responses across trials. Trimmed the data for outliers and using this method of imputation for the missing values will effectively act to reduce intraindividual variability, thereby representing a conservative approach to studying this phenomenon.

Next, because systematic effects due to group, story, or trial may exist in the data, and could possibly confound any group differences later found in intraindividual variability, it was necessary to statistically remove these effects. The data were split into 5 (story) X 3 (statement type) datasets so that the effects of trial (8 trials because there are 8 statements of each type per story) and its interactions could be properly removed, and for each dataset the latencies were regressed onto the age, cognitive status, and trial effects and all their interactions. This procedure produced residual scores that were uncontaminated by group differences in speed of performance, and systematic effects such as practice or effects associated with the stories and their content. These residual (purified) scores were then standardized and converted to T scores. Figure 12 shows the residual T scores for each participant for their latencies to story 1 separated by age group. This figure demonstrates that substantial intraindividual variability among the participants still exists and appears to vary by group even though systematic effects of group have been partialed from the data.

---

8 Leaving the data split into only the 3 statement types regardless of story would mean that some of the 40 trials were separated by minutes (same story/occasion), whereas others were separated by weeks (different story/occasion). Consequently, to effectively remove the trial effects it was necessary to form 15 different datasets so that the trials in each set all pertained to the same story/occasion.
Figure 12.

Mean Latency Residual T-scores for Each Participant as a Function of Age
Results

In order to quantify individual differences in the amount of intraindividual variability in responding to the various statement types, intraindividual standard deviations (ISDs) were computed for each participant for each story. Because the effects of story were not a main focus of this analysis, the 5 ISDs for each statement type were then averaged, resulting in an average ISD for each participant for each of the three statement types.

A 2 (age) X 3 (cognitive status) RM-ANCOVA was performed on the average ISDs, with statement type as the repeated measures variable. As expected, the main effect of age was significant, $F(1, 297) = 35.19, p < .001, \eta^2 = .11$, as the OO ($M = 8.39$) were more inconsistent than the YO ($M = 6.81$). The effect of greater cognitive impairment was expected to result in larger amounts of intraindividual variability, and the MCI-M group ($M = 8.75$) was significantly more inconsistent than the NCI ($M = 7.18$; $M_{\text{diff}} = 1.53, \text{SEM}_{\text{diff}} = 0.31, p < .001$) and the MCI-S groups ($M = 6.87; M_{\text{diff}} = 1.86, \text{SEM}_{\text{diff}} = 0.34, p < .001$), which did not differ. Age and cognitive impairment were expected to interact, resulting in the OO/MCI-M group demonstrating the most inconsistency. Planned contrasts showed no significant differences as a result of cognitive status for the YO. However, for the OO, the MCI-M group was more inconsistent than the NCI ($M_{\text{diff}} = 2.30, \text{SEM}_{\text{diff}} = 0.47, p < .001$) and MCI-S groups ($M_{\text{diff}} = 3.01, \text{SEM}_{\text{diff}} = 0.54, p < .001$), which did not differ. Moreover, the OO were more inconsistent than the YO for the NCI [$F(1, 133) = 13.06, p < .001, \eta^2 = .09$], and

---

9 It could be suggested that the inconsistency analyses could also be focused on accuracy (correct vs. wrong answers). However, because of the high accuracy rate in responding to the story recognition task (True: 88%; False: 93%; Inference: 72%), the likelihood of each participant having a sufficient number of incorrect trials is small.
MCI-M \( [F(1, 77) = 24.28, p < .001, \eta^2 = .24] \) cognitive status groups, but did not significantly differ for the MCI-S group. As expected, Figure 13 shows the greatest effect size was for the MCI-M group, confirming the interactive detrimental effect of increasing age and poorer cognitive status on intraindividual variability.

It was hypothesized that because of the relative greater difficulty in responding to the inference statements, there would be greater inconsistency associated with them. However, the main effect of statement type was not significant. The effect of age was expected to significantly interact with statement type, as the OO adults would be more inconsistent than YO adults on the inference statements, but no significant age differences were expected on true and false items. Unexpectedly, the OO were more inconsistent than the YO for the true \( [F(1, 297) = 16.40, p < .001, \eta^2 = .05] \) and false \( [F(1, 297) = 83.58, p < .001, \eta^2 = .22] \) statements, but there were no age differences in the amount of inconsistency in responding to the inference items (see Figure 14). Moreover, the age difference for the false statements was more than 4 times greater than that for the true statements. However, the amount of inconsistency across the statement types did not differ for either age group.

The Cognitive Status X Statement Type interaction was also hypothesized to be significant. Specifically, decreasing cognitive ability was expected to result in greater amounts of inconsistency on the inference items, but there were no a priori hypotheses concerning inconsistency to the true and false statements items among the various cognitive groups. Posthoc tests (Bonferroni’s correction, \( p < .017 \)) within each group failed to find any significant differences in the amount of intraindividual variability to the statement types. However, the following contrasts were significant (see Figure 15): For
Average ISD as a Function of Age and Cognitive Status

Figure 13
Figure 14

Average ISD as a Function of Age and Statement Type
Figure 15

Average ISD as a Function of Cognitive Status and Statement Type
the true paraphrases, the MCI-M group was more inconsistent than the NCI group \( (M_{\text{diff}} = 1.19, \text{SEM}_\text{diff} = 0.33, p < .001) \); for the false paraphrases, the MCI-M group was more inconsistent than both the NCI \( (M_{\text{diff}} = 2.07, \text{SEM}_\text{diff} = 0.36, p < .001) \) and MCI-S groups \( (M_{\text{diff}} = 2.94, \text{SEM}_\text{diff} = 0.39, p < .001) \), and unexpectedly, the MCI-S group was less inconsistent than the NCI group \( (M_{\text{diff}} = 0.87, \text{SEM}_\text{diff} = 0.35, p < .017) \). Similarly, planned contrasts for the inferences showed the MCI-M group was significantly more inconsistent compared to the NCI \( (M_{\text{diff}} = 1.33, \text{SEM}_\text{diff} = 0.48, p < .01) \) and MCI-S groups \( (M_{\text{diff}} = 2.04, \text{SEM}_\text{diff} = 0.52, p < .001) \). However, as can be seen for both the NCI and MCI-S groups, the greatest mean differences between them and the MCI-M group was on the false statements.

The Age x Cognitive Status x Statement Type interaction was also expected to be significant: Increasing age and cognitive impairment were expected to result in greater inconsistency to the inferences due to their increased difficulty, but there were no a priori hypothesis concerning the true and false items. The 3-way interaction was not significant, and neither were any of the planned contrasts for the inference statements.

**Correlational Analyses**

Finally, we explored the relationships of the measures of intraindividual variability with the overall accuracy, signal detection, and mean latency measures, which can be seen in Table 4. Generally speaking, increased inconsistency was associated with poorer overall accuracy, and more intraindividual variability was also related to less sensitivity to the differences between the statement types. For the criterions, it appeared that a more conservative criterion was positively correlated with increased inconsistency to the true statements, but negatively correlated with inconsistency to the false and
Table 4

Correlations of Intraindividual Variability with Overall Accuracy, Signal Detection Measures, and Mean Latency

<table>
<thead>
<tr>
<th></th>
<th>Average ISD</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>True paraphrases</td>
<td>False paraphrases</td>
<td>Inferences</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total correct to True</td>
<td>-.45**</td>
<td>-.17**</td>
<td>-.12*</td>
<td></td>
</tr>
<tr>
<td>Total correct to False</td>
<td>-.15**</td>
<td>-.52**</td>
<td>-.23**</td>
<td></td>
</tr>
<tr>
<td>Total correct to Inference</td>
<td>.00</td>
<td>-.36**</td>
<td>-.16**</td>
<td></td>
</tr>
<tr>
<td>Signal detection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d'$ (True/False)</td>
<td>-.37**</td>
<td>-.46**</td>
<td>-.23**</td>
<td></td>
</tr>
<tr>
<td>criterion (True/False)</td>
<td>.19**</td>
<td>-.29**</td>
<td>-.11</td>
<td></td>
</tr>
<tr>
<td>$d'$ (True/Inference)</td>
<td>-.25**</td>
<td>-.41**</td>
<td>-.20**</td>
<td></td>
</tr>
<tr>
<td>criterion (True/Inference)</td>
<td>.23**</td>
<td>-.20**</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td>Mean latency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean RT to True</td>
<td>.67**</td>
<td>.50**</td>
<td>.61**</td>
<td></td>
</tr>
<tr>
<td>Mean RT to False</td>
<td>.44**</td>
<td>.75**</td>
<td>.54**</td>
<td></td>
</tr>
<tr>
<td>Mean RT to Inference</td>
<td>.32**</td>
<td>.35**</td>
<td>.49**</td>
<td></td>
</tr>
</tbody>
</table>

*Note. *p < .05. **p < .01.
statements. Also, greater intraindividual variability was associated with slower response time on the story recognition task.
Discussion

Previous research has demonstrated that old-old or cognitively impaired adults are more likely to have poorer recall of stories (Light et al., 1982; Chapman et al., 2002), slower reaction time, and show greater inconsistency on a performance task (Hultsch et al., 2002; Dixon et al., 2004) than younger and more cognitively intact older adults. The present study attempted to build on this research by investigating whether the type of statement tested would also play a role in the outcome of each of these variables. We were specifically interested in the effects of having to recognize an inference that could be drawn from a story. Past research finding that (a) older and MCI adults recall proportionately fewer details related to a story (Stine & Wingfield, 1987; Chapman et al., 2002) and (b) that episodic memory is poorer (Nilsson, 2003), declines faster in old-old age (Colsher & Wallace, 1991), and is one of the most noticeably decreased abilities among those with MCI (Small et al., 2003), formed the rationale for our hypotheses that those with greater old age and those with greater cognitive impairment would have the lowest accuracy of all the participants on the inference statements. Due to their poorer story recall, these individuals would have only limited information to rely on when completing the story recognition task, and thus instead have to trust the plausibility of a test statement, rather than the explicit recall of the statement’s mention in the story. This situation would create a greater likelihood of mistaking an inference, rather than a false statement, as having been part of the presented story. We further expanded this investigation by employing signal detection analyses to obtain a more comprehensive picture of the participants’ discriminability on the task, and in accordance with the above hypothesized recognition process, expected that the OO and MCI adults would
consequently show the least sensitivity to the difference between true and inferred statements. Finally, we extended our hypotheses to the latency and consistency of responding, and expected that the greater relative difficulty of accurately responding to an inference statement would be associated with slower response time and higher levels of intraindividual variability, primarily for the OO and MCI adults.

Summary of Results

For accuracy, the results were generally consistent with our expectations. The magnitude of the relative drop from true/false accuracy to inference accuracy grew with increasing age and cognitive impairment, but this was not the only accuracy deficit associated with these characteristics. The younger and more cognitively intact groups also demonstrated an unexpected relative advantage in the ease of accurately responding to the false compared to true statements. Interestingly, while this advantage was only lowered in the OO adults, it nearly disappeared entirely for the MCI-M adults. As expected, the signal detection sensitivity measures revealed that the true/false discrimination was much easier than the true/inference one. Furthermore, age produced a significant effect as the YO adults were more sensitive to the differences in both discriminations than the OO adults. A similar effect was found for cognitive status, as the NCI group was more sensitive to the true/false and true/inference statement differences than the two MCI groups, and the MCI-S group was more sensitive than the MCI-M group for both statement discriminations. In addition, the combination of older age and poorer cognitive status resulted in a greater drop in sensitivity to the statement differences. The effect of greater sensitivity of the YO adults over the OO adults was nearly five times greater for the MCI-M group than the NCI group in the true/false
discrimination. Likewise in the true/inference comparison, the greatest age effect was for the MCI-M group. The more difficult true/inference discrimination was also associated with more liberal responding compared to the true/false comparison, but group effects were significant for only the true/false distinction, where greater conservativeness in responding was associated with younger age and no or less cognitive impairment.

Unlike the expected nature of many of the accuracy outcomes, the latency and inconsistency results were generally contradictory with our hypotheses. There were no significant overall differences among the statement types in the speed or consistency of responding, but, for each measure, statement type significantly interacted with both age and cognitive status. For latency, the largest age difference was for the false statements, revealing that the YO group maintained an advantage in being able to quickly respond to the false statements, but this disappeared for the OO. Although the pattern of cognitive status differences was similar across all statement types, the mean differences between the MCI-M group and the NCI and MCI-S groups grew successively larger from the easier true statements to the more difficult inference statements, with the largest increase was from the true to false statements. However, for both interactions, there were no differences within each group in responding to the various statement types.

For intraindividual variability, the results from the Age X Statement Type interaction were exactly contrary to our expectations: the OO adults were more inconsistent than the YO adults for the true and false statements, but there was no significant difference for the inference items. Moreover, the age difference for the false statements was more than four times greater than that for the true statements. Our cognitive status hypothesis for inconsistency faired better, as the MCI-M group displayed
more intraindividual variability on the expected inference statements than the two more cognitively intact groups. However, there were also significant cognitive status differences for the true and false statements, where the MCI-M group was more inconsistent than both more cognitively intact groups for the false statements, and than only the NCI group for the true statements. Similar to the effect found as a result of age, the greatest mean differences in inconsistency were on the false statements, but the amount of inconsistency within each group did not differ across the statement types for both interactions.

**Proposed Recognition Process: Recall-to-Reject and Plausibility Strategies**

Although the results from the accuracy analyses were generally consistent with our hypotheses, the latency and inconsistency results turned out quite different than expected. Consequently, the plausibility recognition strategy that drove our hypotheses initially appeared unable to fully explain our discrepant results. However, after considerable re-evaluation we realized that by adding another recognition strategy that may have also been used during the task, the expected reliance on the plausibility of items did in fact adequately explain the present findings.

We propose that the participants may have first attempted to use a recall-to-reject strategy (e.g., Rotello, Macmillan, & Van Tassel, 2000) for the recognition task, whereby any piece of recalled information inconsistent with the test statement would cause the statement to be responded to negatively. For example, if a story had only contained information about an older woman, any mention of an older man in the test would be inconsistent with the information recalled from the presented story, regardless of what other information was included in the rest of the test statement. This may have been the
participants’ first response strategy because they knew that two of the three statements types in the recognition test contained ideas not presented in the story, and therefore were to be negatively responded to. This knowledge may have primed the participants to employ the most effective response strategy which would take advantage of these odds and result in quick and successful rejections of a false or inferred statement. However, in situations where this strategy failed (i.e., the strategy was taking too long, they could not find an inconsistent event, or they simply could not remember any other details from the story), participants may have then switched to the plausibility strategy, where they relied on whether the statement was consistent with what they could remember about the main ideas in the story. Given that the OO and MCI adults have relatively poorer memory, they may have switched to this strategy more often, causing more accuracy errors, revealing less sensitivity to the statement differences, and more liberal responding. We suggest that the OO and MCI adults used the plausibility strategy more often, rather than perhaps relying solely on the plausibility strategy, because in the latter case they would have not been able to maintain their relatively high overall accuracy rate of approximately 80%, as every inferred statement would have been responded to as true and therefore incorrect. Similarly, the fact that the OO and MCI-M adults were disadvantaged on even the relatively simpler true and false statements, which would have resulted in 100% accuracy with the plausibility strategy alone, adds support that more than just the plausibility strategy was involved.

**Evidence in Accordance with the Proposed Recognition Process**

The above theory for how the recognition task may have been approached is supported by numerous results. First, we found that the OO and MCI-M groups
experienced a greater drop in inference accuracy compared to their own performance on the true and false statements. This is what one would expect if these groups' relatively poorer memory reduced the effectiveness of the recall-to-reject strategy, subsequently causing them to switch strategies and rely on the plausibility of a statement more often than the other groups, and resulting in the wrong assumption that the inference statements were presented in the story. Furthermore, the expected consequence of using this strategy would be less sensitivity to the differences between the true and inferred statements, which is exactly what was found for the OO and MCI-M groups.

Even the unexpected relative advantage in responding accurately to the false statements compared to the true statements can now be adequately explained with the addition of the recall-to-reject recognition strategy. This strategy would predict that the incorrectness of the false statements would be more readily apparent because they only require one piece of inconsistent information to be deemed false, compared to having to confirm (or fail to disconfirm) that each piece of information in the true statements was indeed true. For example, in the story which describes a couple's camping vacation (see Appendix), one false recognition statement is, “another older couple invited them over for dinner tonight.” However, a significant portion of the story is devoted to describing what the couple has already prepared or will be preparing for dinner that evening. An individual would thus only have to recall one or two of the food preparation details to logically deduce that the couple would not be preparing food if they were going out for dinner. In contrast, for the true statement, “after bird-watching they return for a hot breakfast,” each piece of that statement has to be verified (i.e., Did they go bird-watching?; Did they have breakfast afterwards?; Was it a hot breakfast?) or there must be
no inconsistent information recalled in order to affirmatively respond to that test statement using the recall-to-reject strategy. Furthermore, the diminished advantage of the false over the true statements for the OO and MCI-M adults is consistent with this theory. If episodic memory decreases with age (Hultsch, Hertzog, Small, McDonald-Miszczak, & Dixon, 1992) and with greater cognitive impairment (Nilsson, 2003), an individual would be able to recall fewer pieces of information from the story, thus reducing the wealth of items they have to call on to accurately complete the recognition task. Fewer pieces of information would lead to a smaller likelihood of being able to reject a false statement, and consequently less sensitivity to make relatively simpler true/false discriminations, as was found for the OO and MCI-M adults.

In regards to latency, both the recall-to-reject and plausibility recognition strategies are needed to fully account for the results. To clearly illustrate this point, we will attempt to explain the latency results with each recognition strategy alone and note where each strategy's predictions are not consistent with the present findings. We will first focus on only the plausibility recognition strategy. We now realize that we incorrectly interpreted how the plausibility strategy would impact response time, and that the failure to find a significant statement type effect for latency is actually consistent with the hypothesized use of the plausibility strategy. If, because of declining memory ability, older adults have to rely on what is plausible with the main ideas of the story when responding to the recognition task, both the true and inferred statements would produce the same ease of responding and consequent response time. However, the plausibility strategy would predict that the false statements would be easily and quickly recognized as being inconsistent with what could be recalled from the main ideas of the story. Despite
this, there is no reason to believe that the false statements would be rejected faster than the true or inferred statements would be accepted, and this is consistent with the lack of a significant statement type effect. However, the theory that the participants used only the plausibility strategy begins to fail when we note that the greatest effects of the Age X Statement Type and Cognitive Status X Statement Type interactions involved the false and inference statements respectively. Therefore, there must have been more to the recognition process than simply relying on what was in accordance with the main ideas of the story.

The recall-to-reject strategy would have the different expectation that the responses to the false statements would be much faster than to the true or inferred statements. As explained earlier, it would have been much easier to find one piece of information inconsistent with the test statement if that statement were false, compared to if it were true or inferred. For example, using this strategy for an inferred statement would have had participants attempting to recall a piece of information inconsistent with the inferred statement and have difficulty finding one. The same result would have occurred when the recall-to-reject strategy was used for a true statement: the participant would have searched for information that was inconsistent with the test statement, but would find that each recalled piece of information instead ended up being consistent with the test statement. Consequently, the search process for both the true and inferred

---

10 Research on choice-reaction time has attempted to answer the question regarding whether errors are responded to faster, slower than, or the same as the corresponding correct responses. The majority of the results support the following pattern: Errors are faster than corresponding correct responses when discrimination is easy and speed is emphasized, but for signals that are difficult to discriminate and when accuracy is emphasized, errors are slower than corresponding correct responses (Luce, 1986). Although discrimination using the plausibility strategy may be considered easy, neither speed nor accuracy was emphasized when the present task was performed. Due to the fact that both criteria (i.e., level of discrimination, speed/accuracy) are needed in order to make a reasonable hypothesis concerning whether response times for errors differ from that for correct responses, we can only assume that the false and the true/inference statements would not differ in their response time under the plausibility strategy.
statements would have taken longer to carry out, causing the RT on the true and inferred statements to be relatively slower than on the false statements. This would explain why the greatest effect of the Age X Statement Type interaction was found for the false statements, but is inconsistent with the Cognitive Status X Statement Type interaction and the lack of a main effect of statement type for latency. The anomalies that remain in each strategy's explanation of the current results reinforce the idea that both recognition strategies are needed to adequately explain the present results.

Finally, the finding that both the OO and MCI-M groups demonstrated the greatest increase in inconsistency for the false statements is also in accordance with the proposed recognition process. Within each statement type, there were statements that were simple and therefore quicker to respond to, and also some more difficult items which required a longer recognition time. It may be the case that, relatively speaking, the YO, NCI, and MCI-S groups were able to consistently use the recall-to-reject strategy and rely on it to make accurate recognition decisions for the false statements. Their greater memory ability would have given them a wider base of information that they could draw on to reject the false statements, allowing them to be generally consistent from the easier to the more difficult false items. Although the OO and MCI-M groups may have also used the recall-to-reject strategy, they have a smaller number of details to base their decision on, resulting in the more difficult items requiring a longer RT, and thus increasing their response inconsistency.

In summary, the ability to efficiently use the recall-to-reject process to a greater extent aided the recognition performance of the younger and more cognitively intact groups. It allowed them to be more accurate overall, and particularly on the inferences,
demonstrate a false versus true statement accuracy and speed advantage, have greater
sensitivity to the statement differences, exert a more conservative criterion of responding
(for true/false), and be more consistent in their responding to the false statements. On the
other hand, the older and more cognitively impaired adults could not rely on this process
to the same extent due to their poorer memory, and consequently showed poorer overall
accuracy, more mistakes when responding to the inferences, and no relative advantage in
accuracy or speed on the false compared to true statements. They also showed less
sensitivity to the statement differences, greater inconsistency in responding to the false
statements, and plausibly due to less confidence in relying on their memory, more
liberality in responding (for true/false).

Inconsistencies with the Proposed Recognition Process

Despite this consistent pattern of results, not all of the present results fit perfectly
with our proposed theory. One such example was the lack of a main effect of statement
type for mean latency. As mentioned earlier, the recall-to-reject theory would predict
that the false statements would be responded to the fastest, while the plausibility strategy
would predict no differences in RT. However, the combination of the two recognition
strategies might predict the following: During the use of the recall-to-reject strategy, only
the false statements would have been given the best probability of being correctly
responded to. The true and inference statements on the other hand would have been
relatively more difficult when using the recall-to-reject strategy, and consequently would
have had a higher likelihood of the participant switching strategies to the easier
plausibility focus. Therefore, one would expect that switching to the plausibility strategy
would have consequent time costs, and as a result the false statements would elicit faster
response times. The failure to find this effect is inconsistent with our proposed theory, but we do offer a speculative explanation of the incongruent result: every single true and inferred statement did not likely elicit the same recognition strategy; some items would have been easier, and therefore feasible to complete with the recall-to-reject strategy; and others would have been more difficult, requiring the participant to switch strategies (of course, the point at which a participant would switch was dependent upon their age and cognitive status). We propose that the high likelihood that each item of each statement type was not completed with the same recognition strategy averaged out RT differences, resulting in the non-significant main effect of statement type in RT.

Finally, there were pieces of evidence which the proposed theory cannot reasonably explain, such as the group interactions with statement type for mean latency and intraindividual variability. The magnitude increases in inconsistency from the MCI-M group to the NCI group were similar for both the true and inferred statements. Although this finding is consistent with the hypothesis that both true and inferred statements were approached in the same fashion, the fact that the differences in RT between the groups progressively increased from the true to inference statements suggests otherwise: We would not expect RT to true and inferred statements to differ if the same recognition strategy was being applied. Conversely for age, there were differences in inconsistency when responding to the true but not inferred statements, but the size of this age difference in mean latency was similar for the true and inferred statements. Overall correlations among the latency measures showed that for each statement type, longer RTs were associated with greater inconsistency, but the above
interactions do not seem to fit this pattern, nor can they be reasonably expected from the proposed theory.

Despite this inconsistency, there are a few instances where the discrepancies involved differences between the age-related outcomes and the cognitive status-related outcomes, suggesting that the theory disagreements may instead involve higher-order group differences. For example, the MCI-M group was more inconsistent than the NCI group across all statement types, while the OO group was more inconsistent than the YO group for only the true and false statements. As well, while the false statement advantage in overall accuracy was only lowered in OO adults, it nearly disappeared for the MCI-M adults. Despite these differences, in each of the examples presented, the MCI-M group demonstrated greater decreased performance relative to their comparative groups than did the OO group to the YO group. The very fact that these same general declines were not seen as a result of age acts as evidence that the changes associated with MCI are often more severe. Furthermore, our results show that while MCI and older age often lead to similar deficits or declines in cognitive abilities, the MCI-M classification is more detrimental to cognitive performance than being of old-old age. These examples illustrate that we cannot expect all results to fall perfectly into place with our proposed recognition process, as the group effects and interactions retain significant influence over the outcomes.

Overall, the results from the present study are well explained by the combination of the recall-to-reject and plausibility recognition strategies. We will now turn to other issues related to our results that are independent of the recognition process and suggest
applied conclusions and future directions encouraged from the present study, as well as comment on any limitations attached to these findings.

**Differences as a Result of the Severity of Mild Cognitive Impairment**

Given that our classification for cognitive status was designed around a continuum from no cognitive impairment to greatest cognitive impairment, one might expect that the NCI group would consistently outperform the MCI groups, and the MCI-S group would perform better than the MCI-M group. However, the pattern of results for cognitive status did not consistently adhere to this expectation. Although the NCI and MCI-S groups significantly outperformed the MCI-M group in the majority of instances, the MCI-S group rarely significantly differed from the NCI group.

We can first conclude that the greater performance of the NCI group, and also on occasion the greater performance of the MCI-S group compared to the MCI-M group verifies that our classification criteria did indeed identify those with lowered cognitive abilities (i.e., MCI-M). Both the NCI and MCI-S groups were more accurate on the false and inference statements, more sensitive to the statement type differences, faster in responding, and showed less intraindividual variability than the MCI-M group.

Next, our conclusion concerning the MCI-S group can logically follow two different paths: First, due to the large number of similarities between the NCI and MCI-S groups, the two different classifications may not represent a stable difference in cognitive status. For example, the MCI-S group was statistically identical to the NCI group in accuracy to the true, false and inference statements, mean latency, and intraindividual variability. Furthermore, given the high overall accuracy of both groups, it remains a possibility that the conversion of the accuracy data into the d' measures artificially
inflated any small differences between the groups, leading to the potentially erroneous conclusion that the MCI-S adults were less sensitive to the statement differences than the NCI adults. These similarities can lead us to conclude that only the multiple distinction of MCI in our classification may be valid, and that our MCI-S group may be cognitively identical to our NCI group. Garrett, MacDonald, and Dixon (2004) have recently showed that the proportion of adults who are classified with Subclinical Cognitive Impairment (SCI) using the greater than 1 SD formula positively increases with the number of tasks used as a classification base, and that the proportion of SCI classifications varies by the cognitive domain(s) and task(s) used. When one considers that most individuals have relative strengths and weaknesses in different areas of cognitive functioning, there is a high likelihood of being greater than 1 SD below age and education matched peers on any one cognitive task. It may be that performing below your peers on two or more cognitive tasks (i.e., the criteria for MCI-M) is a more reliable criterion, and the comparative accuracy of this classification is apparent given the consistent differences between the MCI-M and NCI groups. The lack of significant differences between the two more cognitively intact groups and the inconsistent pattern for those few differences that were significant suggests the reasonable conclusion the MCI-S and NCI groups may not differ in cognitive status after all.

The other possible explanation of the MCI-S results is that because we found even a few significant differences between the NCI and MCI-S groups, differences in cognitive status do indeed exist between the two groups. The MCI-S group had poorer overall accuracy than the NCI group and was less sensitive to both the true/false and true/inference statement comparisons. However, as noted earlier, this difference in
sensitivity must be interpreted with caution. In addition, for the following measures the NCI group outperformed the MCI-M group, but the MCI-S group did not: the relative drop in accuracy performance on the inferences, accuracy to the true statements, and the amount of inconsistency to the true statements. The differences among the MCI groups and between the MCI-S and NCI groups support the conclusion that our MCI-S classification does identify a unique group of adults who are not as cognitively impaired as the MCI-M adults, but not as cognitively intact as our NCI adults. Thus this perspective offers the conclusion that even within the classification of MCI there exists progressing degrees of impairment. However, our results suggest that each progressing point along this MCI continuum does not necessarily result in poorer overall performance than the previous gradation, but rather includes selective declines on specific cognitive tasks.

If the second conclusion concerning our MCI-S group is true, then the possibility that task specific differences exist raises the new issue that identifying individuals with varying degrees of MCI may be especially difficult. There is no universally agreed upon way to classify MCI (Palmer, Fratiglioni, & Winblad, 2003), and most studies' classification schemes depend on the tasks chosen by, or if ad-hoc, available to the researcher. As a result, the sensitivity and stability of the various MCI classification criteria is variable, and similarly, whether those individuals classified as MCI eventually progress to dementia varies from study to study. Therefore, it already appears to be difficult enough to accurately classify MCI, never mind being able to pinpoint individuals with varying degrees or types of the condition. Moreover, the suggested possibility that for these degrees of MCI, the observed differences are not necessarily in line with the
assumption that a lesser stage of MCI simply has less decline, but rather are unique to the task in question, makes eventual accurate diagnosis of these degrees appear near impossible.

Unfortunately, we cannot endorse one conclusion over the other because of the limited available knowledge on MCI. However, we can conclude that the classification of MCI-M does indeed identify a unique group of normally functioning older individuals with impaired, if yet non-clinical, cognitive ability.

Differences in Criterion for Responding

Although the result that the YO adults were more conservative than the OO may be contrary to popular belief, past studies have failed to find a consistent conservative or liberal age effect on response bias (Marquié & Baracat, 2000). However, the NCI and MCI-S groups were also more conservative than the MCI-M group, and the present findings might be due to the OO and MCI-M adults’ reliance on the plausibility strategy. Budson, Daffner, Desikan, and Schacter (2000) found that AD patients had a consistently more liberal criterion in responding to a word recognition task than healthy older adults. They suggested this difference in responding was because the older adults were still able to use item-specific recollection for the recognition task, and could consequently employ a stricter criterion, while the AD adults had to rely on a gist representation of the studied words. These results support our findings that the MCI-M group was the most liberal of all groups on the true/false discrimination, and are in line with the hypothesized reliance of older and MCI-M adults on what is plausible with the gist of the story.

Although this theory is in accordance with the significant group differences seen in the criterion of responding for the true/false analysis, there were no group differences
for the true/inference criterion. In addition to limited power to detect true/inference criterion group differences (.40), the differential findings of the true/false and true/inference discriminations may revolve around the relative difficulty of the two comparisons. As noted earlier, the true and false statements were easy to discern relative to the discrimination between the true and inference statements, and this was reflected by the significant drop in sensitivity for the true/inference analysis. This conclusion can also be drawn from Figures 8 and 9 which illustrate the significant shift in criterion from a conservative (true/false) to a liberal bias in responding (true/inference). Therefore, even the initially more conservative YO, NCI and MCI-S groups had to switch their criterion to a liberal response bias, and despite their greater capacity to directly recall story ideas, they too may have been forced to occasionally use a gist or plausibility strategy.

Effects of Education

The sample used in the present study was highly educated ($M = 15.07$ years) relative to the average aging population. This factor makes our findings all the more powerful, because one would expect that any advantages associated with greater total years of education would conceal any noticeable age or cognitive status effects. In fact, total years of education significantly correlated with accuracy to the false ($r = .33$) and inference ($r = .28$) statements ($p < .001$), in that more education lead to increased accuracy to these two statement types. However, education was not significantly correlated to the true statements. It may be the case that more education lead to better strategy use in being able to differentiate what was actually said in the story from these distractors. The plausibility of this hypothesis is echoed in the correlations of education with the signal detection sensitivity measures. Education was significantly correlated
with the $d'$ (True/False) measure, $r = .27, p < .001$, indicating that the more education one had, the better they were at differentiating between the true and false statements. Interestingly, higher education provided nearly the exact same advantage for the more difficult true and inference comparison, $r = .28, p < .001$, suggesting that while more education aided participants in their accuracy and discriminability, its aid did not linearly increase with the difficulty of the discrimination task. The correlation between more education and greater sensitivity is consistent with previous studies (Marquie & Baracat, 2000), and demonstrates the importance of having covaried education in all of our analyses.

Education was also a significant correlate of the criterion (True/False), $r = .28, p < .001$, showing that the more education one had, the less risky they were in their discrimination. A similar effect of higher education was found for the criterion (True/Inference), $r = .23, p < .001$, as it lead to greater conservatism in the true and inference comparison. Marquie and Baracat (2000) found a significant age effect on the response criterion only in their most educated group ($\geq 14$ years), where the older the participant, the more conservative their response bias. Our sample was similarly highly educated ($M = 15.07$ years), perhaps explaining its correlation to the criterions.

Finally, education appeared to differentially correlate with the latency and intraindividual variability to the statement types. The total years of education completed was significantly correlated with mean latency in responding to the true ($r = -.13, p < .05$) and false ($r = -.23, p < .001$) statements, whereby the higher education level attained, the faster one was able to respond. However, education was significantly correlated with the amount of inconsistency to the inference ($r = -.12, p < .05$) and false ($r = -.17, p < .01$)
statements, as more years of education meant less intraindividual variability within responding to these statement types. There is no clear rationale for this differential pattern. Overall, the many significant correlations of education with the outcome variables reiterate the importance of having used education as a covariate in these analyses, and the ability to even then find significant effects makes these results all the more robust.

Inferences and Intraindividual Variability

At first glance, the failure to find the hypothesized greater inconsistency for the more difficult inference items demonstrates that more difficult items within the same task may not necessarily cause relatively greater intraindividual variability. However, the types of statements may have been too similar to one another to result in significantly different levels of inconsistency. Furthermore, increased inconsistency is associated with slower reaction time, but there were also not any significant statement type effects in mean RT. Previous studies have found intraindividual variability differences between verbal (semantic and lexical decision) and non-verbal tasks (SRT and CRT; Hultsch et al., 2002), and executive (1-back RT) and non-executive tasks (SRT and CRT; Dixon et al., 2004) but to date, no study has found differences within the same task. Although there were no main effects of statement type, significant differences in the amount of inconsistency within the different statement types were found when group factors were taken into account. Increasing age resulted in more intraindividual variability for the true and false statements, and poorer cognitive status was associated with greater inconsistency for all 3 statement types. In both cases the largest effects were unexpectedly found on the false statements, and although they were not the most difficult
to recognize, given the proposed recognition process, they would have elicited the
greatest inconsistency in responding.

Variations in Results Depending on Statistical Approach

The application of both the overall accuracy and signal detection analyses to the
data revealed subtle differences of story recognition in older adults that otherwise would
have been overlooked if only one statistical technique was used. For example, although
the overall accuracy analyses failed to find any significant group interactions, the signal
detection measures revealed significant Age X Cognitive Status effects on the basis of
sensitivity to the differences between the statement types. Furthermore, we found that
how data are analyzed can factor into whether differences among cognitive status
classifications are revealed. Employing both accuracy techniques showed that although
the MCI-S group did not differ from the NCI group on the overall accuracy measures to
each statement type, the MCI-S group did have poorer sensitivity to the statement
differences. The present study demonstrated that particular attention must be paid to
which statistical technique is chosen and that different conclusions may result from
different statistical applications.

Applied Conclusions and Future Directions

We can conclude that old-old and/or MCI adults may be less able to recognize
that something plausible and consistent with an event may not have actually occurred.
These adults are more likely to confuse an idea that could be inferred from a story as
actually having been part of the story, which leaves them vulnerable to misinterpretation
and has a wide range of potential implications. Consider the following example: An
older adult hears of an exciting new health discovery on the news, and when they next
see their family, he or she recalls this exciting story. However, when doing so, they inadvertently recall some inferences as fact in their reiteration of the story because the inferences seem likely to have also occurred. While the addition of these assumptions is not harmful in this example, consider the potential social implications: If others notice that an older individual frequently embellishes stories (in this example, they are able to recall that these added details were not actually included in the news report), they may begin to form incorrect impressions of the older adult, ranging from the idea that the older adult is credulous, to declining in cognitive ability, to simply craving attention that they hope to attain by adding these untrue details. The implications of this potential vulnerability can also be extended to everyday problems such as those concerning health care or finances, where relying on what is plausible with what one might expect has a much greater likelihood of having serious consequences. Furthermore, the possibility exists that these types of individuals may be especially vulnerable to situations which rely on assuming that certain inferences are true (e.g., scams). These types of situations want individuals to not ask clarifying questions and simply make plausible inferences such as the legality of the proposal, the quality of the supporting research, and even that scientific research has been done.

The significant real-world implications of these findings highlight the need to examine these results in a more applied context. One could expand the same story recognition task to a more common real life situation, such as a description of a health problem and suggested treatment plan, or a desired financial situation and possible banking options. Although we generalized the current results across five different stories, constructing this new task would place it into the everyday problem solving domain,
allow us to have some of the necessary external validity needed to further support the conclusions from the present study. There are obvious ethical limitations in testing the conclusion that these adults are more vulnerable to scams, in addition to the fact that one’s potential vulnerability to a scam would most likely include a number of other factors, such as personality (e.g., riskiness) or social (e.g., loneliness) variables. However, our results are consistent with the conclusion that OO and/or MCI adults may need to be particularly cautious of these situations, and research and public dissemination of this vulnerability need to be explored further.

It is important to note that the present results also showed that the younger and more cognitively intact older adults were also relatively less able to reject inference statements compared to true or false statements. Thus, these adults may also be vulnerable to the described situations, even if less so than OO and MCI-M adults. At the other end of the spectrum, those with both older age and increased cognitive impairment demonstrated the worst sensitivity, speed, and consistency of responding of all groups, indicating that this group may need to be especially cautious. Finally, older age and poorer cognitive status were associated with reduced recognition to even the easier true and false statements, suggesting that declines in episodic memory that occur with age and cognitive impairment can impact even relatively simple cognitive tasks.

Although our stance on the present results was to support the notion that older adults recall proportionately fewer details due to memory difficulties rather than developmental changes as proposed by Adams (1991), both cases predict that older adults have fewer details available to them. While the present study cannot explicitly provide any conclusions to that debate, the developmental hypothesis would have difficulty
accounting for the similar deficits in ability found between those of greater older age and those with cognitive impairment.

We have two potential and competing conclusions which can be drawn from the current MCI-S results: The first is that given the large number of similarities between the NCI and MCI-S groups, these two groups did not really differ in cognitive status, and thus the classification scheme used to divide them is invalid; The second conclusion states that the fact that the NCI and MCI-S groups did significantly differ on a few outcome measures demonstrates that a continuum of MCI impairment exists, and that declines along this continuum are particularly task specific. At the present time and given the current information available concerning MCI, we unfortunately cannot support one conclusion over the other. Regardless of which conclusion is correct, future research needs to address the core issue of determining the most appropriate cognitive domains, tasks, and cut-offs to discern those with MCI from healthy older adults. Longitudinal research is also required to confirm the validity of our classification criteria by observing whether those individuals classified as MCI do go on to develop dementia.

One of the reasons for including cognitive status in this study was to learn more about how differences in story recognition may differentiate those with MCI from healthy older adults. Chapman et al. (2002) found that relative to healthy adults, the recall of both the main ideas and details of a story diminish in AD, while in MCI, only the memory for details is significantly poorer. These findings suggested that the preservation of lower level propositions of a story, and therefore inferences, may uniquely differentiate MCI adults from NCI adults on a story recognition task. In fact, we found effects of cognitive status for all statement types, but the greatest accuracy differences did
indeed involve the inferences. Our next step would be to compare AD and MCI adults on this task to determine if recognizing inferences can distinguish these two clinical classifications.

Another direction of future research surrounds the question of what intraindividual variability truly represents. Our present findings are in line with previous research that an individual appears to become more inconsistent on a performance task with increasing age (e.g., Williams et al., 2003), more severe MCI (Dixon et al., 2004), and the interactive effect of these two factors (Dixon et al., 2004). Although we can confidently conclude from the results of the present and past studies that inconsistency is a stable, latent trait of an individual, we still cannot definitively say what inconsistency actually is or what it is caused by. The current assumption that intraindividual variability is a marker of neurological integrity fits the inconsistency research very well, and also draws support from the present study, which found that increased inconsistency was generally associated with poorer overall accuracy and less sensitivity to differences between the statement types. However, more research is needed to confirm this theory and determine if and why inconsistency may be the by-product of neurological decline.

Limitations

Although the present study expands our understanding of story recognition among older adults, there are a number of limitations that should be considered. First, the lack of universal MCI criteria is an issue, as our classification scheme may not be the same one employed by another research group. In addition, although the significant results of this study were typically related to those individuals with MCI-M but not MCI-S, very few studies further divide their MCI group into degrees of impairment, limiting the
replicability of our findings. The current sample was also highly educated, and although one could make the argument that this limits the generalizability of the study, it in fact makes our results stronger because significant differences were still found as a result of age and cognitive status.

Of course the greatest limitation of this study involves the proposed recognition process. Our proposal for how the recognition process may have been carried out for this task is consistent with the vast majority of the results, but it must be acknowledged that this proposal was developed after the fact, and is quite speculative. The present results cannot explicitly verify that the OO and MCI-M adults did in fact switch to the plausibility strategy sooner and more frequently than their younger and more cognitively intact peers, and consequently the proposed recognition process can act only a tentative hypothesis.

**Overall Conclusions**

The present study expanded the current story recognition and inference literature by investigating age differences within the older age range, differences as a result of cognitive status, and extending the focus of the investigation into response consistency. We were able to confirm our hypotheses that increasing older age and greater cognitive impairment result in poorer ability to accurately recognize inferences, and reduce sensitivity when discriminating between true, false, and inferred statements. The amount of intraindividual variability demonstrated by an individual was positively correlated with their age and cognitive impairment, but surprisingly was not uniquely affected by the type of statement one was responding to. There were however, significant Group X
Statement Type interactions, which revealed that the greatest inconsistency involved the false, rather than the hypothesized inferred statements.

We propose that two different recognition strategies were involved when participants were completing the story recognition task, and that their episodic memory ability played a defining role in the efficiency and consequent frequency of use of the two strategies. Although our hypothesized recognition process is only speculative, the results from the current study provide substantial support in favour of its plausibility. Old-old and MCI adults appear to be more susceptible to interpret something that is plausible and consistent with an event as having actually occurred, and are even more likely to misinterpret a false statement as true and vice versa. These results have real life implications which require further investigation, in addition to highlighting the need to inform older adults and develop strategies that could be used to combat this potential vulnerability.
References


Appendix

Story 3: Camping

William and Mildred are camping again this summer in Northern Ontario. Each year they drive from their home near Toronto, Ontario, to their favourite woodland resort. They have been spending their summers here for twenty-three years. They always rent the same cabin. It is sunny in the morning and shady in the afternoon. While William goes fishing in one of the nearby lakes or streams, Mildred reads magazines or novels, or works on her knitting. Sometimes they get up before dawn and go bird-watching together. Afterwards, they go back to their cabin for a hot breakfast. All the park rangers know them by name. They enjoy getting to know the families that visit the park for the first time. They also enjoy getting reacquainted with the families that come to the park regularly. This summer there is another older couple from Manitoba visiting the park, too. William and Mildred have invited them over to dinner tonight. William caught seven pan-sized trout this morning and has already cleaned and cooled them. Together with the four he put on ice yesterday, there should be enough fish for everyone. He just finished peeling 16 small potatoes and slicing one very strong onion. Mildred will prepare her special home fries. She just finished shucking eight ears of sweet corn. The other couple, the Wilsons, promised to bring something for dessert. William and Mildred treasure the opportunity to have feasts like this with new or old friends. It is one thing that brings them back to this resort every year. They also like the clean fresh air, the warm relaxed atmosphere, and the chance to enjoy the outdoors. They dearly love their home in Toronto but every year they look forward to their Northern Ontario vacation. It is a place where all their desires are fulfilled.
Appendix (con’t)

Story 3: Questions

1 = True paraphrase; 2 = False paraphrase; 3 = Inference

3 As children, William and Mildred learned to love camping.

3 William and Mildred have always preferred a simple life.

3 After their vacation they are relaxed and ready to start again.

2 Both William and Mildred go fishing often.

2 William and Mildred sometimes resent newcomers.

2 Mildred is preparing a dessert.

3 They enjoy showing first-time visitors around the camp.

1 After bird-watching they return for a hot breakfast.

1 William and Mildred are camping again in Northern Ontario.

2 Each summer they try a different cabin.

1 William caught seven trout this morning.

1 All the local park rangers know William and Mildred.

2 Another older couple invited them over for dinner tonight.

2 They will also cook hamburgers, because there won't be enough fish.

3 William and Mildred are members of the Audubon Society.

1 They have gone camping in the same place for twenty-three years.

2 They sometimes complain about increasing air pollution in the area.

1 Every year they meet both first-time and regular visitors.

1 Mildred is preparing home fries.

1 They treasure the opportunity to have feasts with old friends.
Appendix (con’t)

3 William is more skilled as a fisherman than Mildred.

2 To go camping they first fly to Quebec City.

3 Mildred has been a homemaker since she married William.

3 William and Mildred have no desire for long or exotic vacations.