# Confidence Based Calibration and the Detection of Early Cognitive Loss in Probable and Possible Alzheimer's Disease Sufferers

by

Joanne E. Minns

A thesis submitted to

The faculty of Graduate Studies and Research

in partial fulfillment of the requirements for

the degree of

Master of Arts.

Department of Psychology

Carleton University

Ottawa, Ontario

May 5, 2002

The undersigned recommend to the

Faculty of Graduate Studies and Research
Acceptance of the thesis:

**"Confidence Based Calibration and the Detection of Early Cognitive Loss in Probable and Possible Alzheimer's Disease Sufferers"**

submitted by

Joanne E. Minns

in partial fulfillment of the requirements for

the degree of Master of Arts

Chair

External Examiner

Thesis Supervisor

## Abstract

Examined is the realism of confidence in the memory and perceptual domains for two groups of seniors (i.e., seniors diagnosed with Alzheimer's Disease (AD) and seniors who have not been diagnosed with AD). For the Choice Reaction Time, Stroop and Population of Cities in Canada tasks, participants were at least 13 seniors (age 55+) suffering from AD. On the Inspection Time task, this group numbered 9. At least 49 seniors who had not been diagnosed with AD participated in all four tasks. Performance of these groups was compared on both a memory task (Cities in Canada) and a perceptual task (Inspection Time). As expected, difficulty effects were evident in both sensory and perceptual domains at all levels of stimulus difficulty with improved performance on easy items compared to hard items. No group differences were found on any of the calibration indices on the Cities in Canada memory task. However, contrary to expectations, group differences on calibration indices were also not found on the perceptual task of Inspection Time. Based on resolution results, it is interesting that, although performing more poorly than the Non-AD seniors on this task, the AD participants were aware of their inabilities to perform at the most difficult levels whereas their Non-AD counterparts did not show the same awareness. As well, very large differences between the AD and Non-AD groups were found in reaction times on both tasks. Also, seniors diagnosed with AD showed distinctive relative frequencies patterns in their use of the confidence categories as compared to non-AD seniors on both tasks at all levels of difficulty. These findings suggest the potential use of reaction time and the use of calibration indices in the detection of early cognitive loss due to AD.

## Acknowledgments

I dedicate this thesis to Dr. David E. Minns who has provided unending support throughout the unusually long process associated with the achieving of this goal. He has always been my cheerleader, my editor, my financier, my father, my hero. The example he has set for me shall be taken with me throughout my life. Thanks Dad. Also, the support, assistance and incredible knowledge provided by Dr. William Petrusic cannot be overstated. I am hugely appreciative for having had him take on this challenge and for having made it work. Thank you for helping provide the drive and "confidence" when it felt as though I had neither the steam left to make the engine run nor ability to steer the vehicle. Big thanks also go to Derek A. Harrison without whose excellence in SAS based programming, information and words of wisdom I would have been truly lost. I owe you many drinks! Throughout this endeavor, there has been one woman who has offered her support and encouragement as I encountered each of many hurdles. Lynn Giff has been the epitome of kindness and there are not enough words to indicate my gratitude for all she has done for me. It has not gone unnoticed. Thanks to Sherry Lampert at the Pacific Council of Senior Federal Officials for her understanding and support. Also, to Dr. Paddi O'Hara for being her. Her intelligence and strength have been an inspiration. Finally, thanks go to my husband, James, who has provided me with love, support, patience, Smarties, ibuprofen and Kleenex which were needed regularly and often at the most inopportune times. Thanks for encouraging me to stick with it; you have been my pillar of strength.

This research was funded, in part, by a National Science and Engineering Research Grant (NSERC) to William M. Petrusic as well as by a grant from the American Alzheimer's Association to Kathryn Oakley. Thanks also to William Petrusic for the loan of the timing board and response panel and to Centre Elisabeth-Bruyere (SCO Hospital), particularly Dr. D. A. Guzman, Director, Memory Disorder Clinic, for their support and provision of space.

## Table of Contents

## List of Tables

**TABLE** **Page**

## List of Figures

**FIGURE** **Page**

# CONFIDENCE BASED CALIBRATION AND THE DETECTION OF EARLY COGNITIVE LOSS IN ALZHEIMER'S DISEASE

With the proportion of elderly in the general population growing steadily, Alzheimer's disease (AD), diagnosis and treatment, is becoming more of a concern each year. Early stages of AD are difficult to differentiate from cognitive change that results from normal aging, and early diagnosis of AD has crucial implications for the effective treatment of the disease. With current advances in neuroscience and molecular biology, treatments for AD can be expected to be developed that may delay or halt the dementing process but are unlikely to reverse it (Green *et al*, 1997; Small, 1997). For this reason, early detection of AD will be critical in minimizing the damaging effects of AD as advances are made in the area of treatment. Recent research suggests that the area of metacognition (i.e., confidence based calibration) may be helpful in making the distinction between normal aging and cognitive loss found in seniors with AD (Oakley, 1998). In order to directly evaluate the efficacy of metacognition in the detection of early cognitive loss in older adults, first an overview of AD is presented, followed by a discussion of various approaches to the measurement of metacognition (i.e., anosognosia, feeling-of-knowing, and confidence-based calibration). In this context, three models of confidence judgements (subjective probability) are reviewed and several hypotheses are presented.

In 1991 the proportion of the population that was over the age of 65 in Canada was 10.6% and is expected to rise to 14.5% by the year 2011 and to 21.8% by 2031 (Canadian Study on Health and Aging (CSHA), 1994). Of those aged 65 and up, the prevalence rates of AD reported by various researchers ranged from 1% to 15.3%, and

increased dramatically with each decade to 20-47% in those aged 80 and up, and 58% in those aged 95 and over (see *Table 1*). As the segment of the population aged 65 and up is growing rapidly, dementia, particularly AD, is a problem for which finding a solution is becoming increasingly urgent.

| Author | Prevalence Age 65+ | Prevalence Age 80+ | Prevalence Age 95+ |
|---|---|---|---|
| Mortimer, 1983 | 4 | 20 | |
| Evans et al., 1989 | 10 | 47.1* | |
| CSHA, 1994 | 8 | 28.1* | 58 |
| Rocca et al., 1986 | <1 - 10 | | |
| Pfeffer, 1987 | 15 | 35 | |

* Evans et al., and CSHA were based on 85+

*Table 1*. Partial Overview of Prevalence Rates of AD by Age Group (%)

AD is the most common form of dementia, making up 50-80% of dementia cases (Chui, 1989; Evans et al., 1989) and is an irreversible and progressively degenerative brain disease (Alzheimer Society of Canada, 1991). AD is associated with gradual degeneration of the nerve cells of the brain (i.e., senile plaques that occur in large numbers within the cerebral cortex, hippocampus, amygdala and neurofibrillary tangles, debris and shrinkage in the neocortex and hippocampus) resulting in the progressive inability of sufferers to make sense of the outside world and to send messages to their bodies (Mayeux & Schofield, 1994; Selkoe, 1989). Furthermore, most patients with AD

have a slight reduction in brain weight and mild to moderate cerebral cortical atrophy may be present, although many patients show little or no atrophy (Mayeux & Schofield, 1994). There is disagreement in the literature as to which parts of the brain are the most affected by AD. Although positron emission tomography (PET) has shown that the association neocortex, hippocampus, and amygdala are more affected than primary neocortex, it has not been determined whether the parieto-temporal or the frontal regions of the neocortex are more affected (Grady & Rapoport, 1992). All researchers agree, however, that individuals with AD eventually become unable to think clearly, remember, or make decisions (Alzheimer Society of Canada, 1991).

The risk of death has been reported to be substantially increased for individuals diagnosed with AD in either community-based and institutional settings (Evans et al., 1989). Although the cause of the disease is not known, the research literature has identified three possible contributing factors: genetics (e.g., accumulation of amyloid beta-protein; Selkoe, 1989), environmental factors (e.g. exposure to Aluminum; Forbes et al., 1995), and internal factors (e.g., viruses, chemical imbalance, immune system problems; Alzheimer Society of Canada, 1991).

These factors make the assessment of Alzheimer's disease very complex. Currently, the only definitive way of determining the existence of the disease is with a post-mortem examination. Advances have been made to identify antemortem markers of AD but they have not been sufficiently developed to provide reliable and specific discriminators (Prinz & Vitiello, 1989). Consequently, physicians can only make a diagnosis of 'possible' or 'probable' Alzheimer's disease in living patients. In order to make a diagnosis of 'possible' or 'probable' AD, indications of dementia must be

established (i.e., an insidious onset with a generally progressive deteriorating course) as well as reasonable elimination of all other specific causes of the dementia (e.g., stroke, alcohol abuse, head injury). This is done through the use of patient history, physical examination, and laboratory testing (Raskind & Peskind, 1992). A diagnosis of 'probable' AD is suggested for patients between the ages of 40 and 90 who are found to have intellectual decline (e.g., memory, judgement, language, or perceptual problems) that can be documented by neuropsychological tests in the absence of any other disorder that may cause this decline (McKhann *et al.*, 1984). 'Possible' AD is used when a condition is present that may be contributing to the dementia but is not considered to be a causal factor (e.g., stroke; McKhann *et al.*).

The most consistent symptom of intellectual decline in patients with AD is progressive memory loss with deterioration in executive system function. Memory loss in AD is characterized by the inability to learn new facts and to access semantic memory as well as difficulty retaining newly learned information over time (La Rue *et al.*, 1992; Storandt *et al.*, 1984). Memory deficits in AD patients differ from normal seniors in their maintained inability to remember despite cuing. Healthy seniors, on the other hand, benefit substantially from the effects of cueing (Masur *et al.*, 1989; La Rue *et al.*). As the disease progresses, seniors with AD perform more poorly on all memory indices, with scores on recall, delayed recall, recognition, and retrieval from long-term storage distinguishing AD from normal aging with 80% sensitivity (Masur *et al.*). Deficits in memory, particularly for recent events, are reported to precede other impairments in orientation, judgement, problem solving, language and perception (Mayeux & Schofield, 1994; Moss & Albert, 1992; Raskind & Peskind, 1992). These deficits in memory,

however, can be quite subtle because they are often observed to a mild degree in healthy seniors (Giambra *et al.*, 1995). This difference in onset of symptoms is important for the current research as this indicates that differences may also exist in performance abilities between perceptual and memory tasks (i.e., that perceptual abilities may be maintained longer than semantic memory).

Although there are currently no easily applied measures to distinguish early AD from normal aging, such a measure would be of great value to the health care system in order to allow for early intervention and treatment. Such a measure would have to be sensitive to the effects of early cognitive loss, but not to age as the majority of AD sufferers are seniors and one would want to be sure that results are due to cognitive decline and not to normal aging.

Oakley (1998) recently proposed that metacognitive functioning may be sensitive to early cognitive loss in older adults. A review of the literature identified three approaches to the study of metacognition: anosognosia, feeling-of-knowing (FOK) and confidence based calibration. Anosognosia (unawareness of one's own memory loss) is usually determined by the difference score between self-reporting on a memory questionnaire and either an informant-rated memory questionnaire or with performance on objective testing (Mullen *et al.*, 1996). Mullen *et al.* suggest that anosognosia is useful as a potential indicator of early cognitive loss in Alzheimer's disease. Michon *et al.* (1994), however, found that anosognosia in AD appears to be unrelated to the degree of cognitive deterioration, but instead may result from frontal lobe dysfunction. Other researchers have also consistently found that anosognosia was not associated with disease

duration, dementia severity or degree of memory impairment (Feher *et al.*, 1991; Mullen *et al.*, 1996; Reed *et al.*, 1993).

Vasterling *et al.* (1995), on the other hand, suggested that anosognosia may be domain specific with the greatest impairment of awareness occurring in the domains of memory and self-care. These researchers and others noted that anosognosia appears to be associated with more advanced disease and greater dementia severity (Lopez *et al.*, 1994; Migliorelli *et al.*, 1995; Starkstein *et al.* , 1997). Cotrell (1997) further pointed out that patients vary greatly in their degree of insight and areas of self-awareness, further making anosognosia an inappropriate tool in determining the early onset of cognitive loss. No studies were found reviewing anosognosia in the perceptual domain. The inconsistencies in the literature clearly challenge the utility of anosognosia as an index of early cognitive loss.

A search of the literature on metacognition and Alzheimer's disease revealed a small body of research investigating confidence ratings of "feeling-of-knowingness" (FOK) in AD. FOK refers to the subjective assessment of confidence used to investigate the ability to monitor information in memory. The FOK rating is usually determined by asking participants to rate their confidence in the likelihood of future correct recognition of a previously incorrect response (Pappas *et al.*, 1992).

Research findings reveal conflicting results. Pappas *et al.* (1992) evaluated FOK in normal seniors and AD patients on long-term and episodic memory tasks. They found that normal seniors and those with AD were equally accurate in assigning confidence ratings related to the probability for correct recall despite the actual impairment of recall by AD participants. The data from patients with AD did, however, indicate impaired

ability in predicting subsequent recall. The findings further indicated that AD patients showed impaired FOK for knowledge memory but intact awareness of recall veracity from knowledge and episodic memory suggesting a potential uncoupling of long-term memory content from awareness of the veracity of recalls from long-term memory.

Conversely, Lipinska and Bachman (1996; Bachman & Lipinska, 1993) suggested that AD is associated only with deficits in the ability to retrieve information from semantic memory, as indicated by reduced accuracy in both recall and recognition of factual information (i.e., general knowledge questions). Whether this deficit is due to loss of information or to impaired access to that information was not determined. Both of their studies showed increased FOK ratings associated with increased accuracy for both groups. Despite difficulty associated with fact retrieval, however, Bachman and Lipinska found that there was no evidence for defective metacognitive monitoring of general knowledge in participants with AD compared to normal seniors. That is, no difference in confidence judgements was noted between the two groups. Based on Bachman and Lipinska's findings, metacognition (with respect to general knowledge) appears to remain intact despite fact-retrieval problems, contrary to the interpretation by Pappas *et al.* (1992). Unfortunately, the research on FOK is not considered relevant to the current research as it uses gamma correlations. With these types of correlations participants tend to use only one category when making their judgements resulting in the absence of differentiation (i.e., normal seniors differ from those with AD on performance but not on FOK). Because this measure is not clearly sensitive to cognitive loss, its utility in early diagnosis of AD is not considered useful in this regard.

An alternative measure of metacognitive functioning (i.e., confidence based calibration) in normal older adults has recently been explored (Oakley, 1998; Oakley-McKeen & Petrusic, 1995; Oakley-McKeen & Petrusic, 1996). These researchers reported that the use of confidence ratings about the accuracy of one's decisions may prove useful in distinguishing early signs of cognitive loss from normal aging because of differential sensitivity of the three calibration indices (i.e. the calibration index, under/over confidence and normalized resolution).

***Calibration indices***

Typically[1], the probability assessments, $\psi_i$, are partitioned into $J$ categories with $n_j$ occurrences in category $j$ and $n = \sum_{j=1}^{J} n_j$. Letting $\epsilon_{ij}$ denote the occurrence of the event $E$ (a correct response) conditional on the $j$th category, the well known Murphy (1973) partition of the *Mean probability score* can be obtained and is given in Equation 1:

$$\frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(\psi_{ij}-e_{ij})^2 = \frac{1}{n}\sum_{j=1}^{J}n_j(\overline{\psi}_{.j}-\overline{e}_{.j})^2 - \frac{1}{n}\sum_{j=1}^{J}n_j(\overline{e}_{.j}-\overline{e}_{..})^2 + \frac{1}{n}\sum_{i=1}^{n}(e_{ij}-\overline{e}_{..})^2, \quad (1)$$

where $\psi_{ij}$ is the $i$th probability assessment into category j, $e_{ij}$ denotes the occurrence of the $i$th correct response given confidence category $j$, $\overline{\psi}_{.j}$ is the average probability assessment (confidence) in category $j$, $\overline{e}_{.j} = p_j(correct)$, the mean probability of the event occurring (i.e., the probability of a correct response) *conditional* on a probability assessment that falls into the $j$th category, and $\overline{\epsilon}$ the overall probability of a correct response.

*Calibration Index (CI).* The first term to the right of the equals sign in Equation 1, defines the calibration index

$$\frac{1}{n}(\sum_{j=1}^{J} n_j(\overline{\psi}_{.j} - \overline{e}_{.j})^2. \quad (2)$$

The calibration score provides a weighted index of how closely the mean probability assessment in category $j$ matches the obtained empirical probability of the occurrence of event $E$ and varies between an optimal value of 0 (i.e., perfect calibration) and 1 (the worst possible calibration). As an example of the latter case, the judge would have to report absolute certainty on each trial and always be wrong. In practice, however, calibration scores above 0.10 are rarely encountered. The global *calibration-in-the-large* index (Yates, 1990) which is also, more typically, referred to as *Over/Underconfidence (O/U)* is defined below.

*Over/Underconfidence (O/U) Index.* The index of over/underconfidence is defined by the signed difference between the average, overall, confidence rating, $\overline{\psi}_{..} = \overline{conf}$, and the average, overall, proportion correct, $\overline{e}_{..} = p(correct)$; i.e.,

$O/U = \overline{\psi}_{..} - \overline{e}_{..} = \overline{conf} - p(correct)$.

*Resolution Index.* Resolution, or discrimination as it is sometimes called is defined by

$$\frac{1}{n}\sum_{j=1}^{J} n_j(\overline{e}_{.j} - \overline{e}_{..})^2 . \quad (3)$$

The resolution index, based on the variability of conditional probabilities of event occurrence, provides a quantitative index of the ability of the judge to use the $J$ confidence categories to effectively distinguish when the event E occurs and when it does

not. The resolution index is bounded above by the overall variability of the indicator variable; i.e., by

$$Var(e_{ij}) = \bar{e}_{..}(1-\bar{e}_{..}) = \frac{1}{n}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(e_{ij}-\bar{e}_{..})^2 ,$$

which is also the final term in Equation 1 and it is also referred to as "*knowledge*" (Lichtenstein & Fischhoff, 1977). In the two-alternative, forced-choice case, resolution assumes a maximal value of 0.25 when $\bar{e}_{..} = p(correct) = 0.5$ and the judge correctly assesses occurrence and non-occurrence of the event $E$. On the other hand, the lower bound on the resolution score is 0, which denotes a complete inability to use the $J$ assessment (i.e., confidence) categories to differentiate event occurrence and non-occurrence (i.e., correct from incorrect responses in the binary choice context).

*Normalized Resolution.* The final term in Equation 1 defines the upper limit of the resolution index. Consequently, Yaniv *et al.*, 1991 (see also Sharp *et al.*, 1988) have recommended that the raw resolution score be normalized using this term (i.e., $Var\ (e_{ij})$). This normalized resolution index, is thus given by

$$NRI = (\frac{1}{n}\sum_{j=1}^{J} n_j(\bar{e}_{.j}-\bar{e}_{..})^2)/\bar{e}_{..}(1-\bar{e}_{..}) = Resolution/Var(e_{ij}) = \eta^2 ,$$

and is interpretable as the between category portion of the overall variance; it is directly comparable to the $\eta^2$ measure typically encountered in analyses of variance.

*Cognitive deterioration and calibration research*

Oakley (1998) found that the use of these calibration indices may be promising in the detection of early signs of cognitive loss in older adults. More specifically, she determined that these indices, when combined over four perceptual tasks and one

memory task for each individual (thereby creating three global calibration indices) were instructive in teasing apart the effects of age from cognitive loss. That is, normalized resolution was sensitive to the effects of age (i.e., young adults were significantly better resolved than seniors) and the Calibration Index (CI) was sensitive to the effects of early cognitive loss (i.e., seniors who met an *a priori* screening criterion of a mean choice response time of 1 second were distinguished from those seniors who did not meet this criterion). These findings suggest that the CI may be useful in detecting cognitive loss associated with AD although this was not directly tested. Consequently, for the present study, two of the tasks used by Oakley were selected to directly investigate the sensitivity of the CI in differentiating normal seniors from AD patients.

The calibration indices, both CI and Resolution, have also been shown to be sensitive to differences in stimulus difficulty. This is useful in that a patient with AD may perceive the stimulus as more difficult than normal seniors. The "difficulty" effect (Lichtenstein & Fischhoff, 1977) or "hard/easy" effect (Gigerenzer *et al.*, 1991) refers to the relationship between under/over confidence and difficulty. Calibration researchers have noted this effect in both general knowledge and perceptual domains. For example, researchers have noted over-confidence with difficult items and under-confidence with easy items in the general knowledge domain (Nickerson & McGoldrick, 1995) as well as in the perceptual domain (Baranski & Petrusic, 1994, 1995, 1999; Petrusic & Baranski, 1997). Baranski and colleagues further noted that the resolution of confidence was difficulty dependent (i.e., normalized resolution values decreased as the difficulty of the perceptual discrimination increased). The findings indicated overall lower over-

confidence, better resolution (higher resolution values) and improved calibration (lower CI values) for easy items compared to hard items.

Other researchers, however, have not replicated these results in that they have found only under-confidence in the perceptual domain (Dawes, 1980; Keren, 1988; Olsson & Winman, 1996). Bjorkman *et al.* (1993) also failed to find over-confidence in the perceptual domain but did not take in to account the relative lack of difficulty found in the perceptual task versus the general knowledge task used (Oakley, 1998).

Recently, Petrusic and colleagues examined the robustness of the "hard/easy" effect across domains. Aubin and Petrusic (1997) presented evidence for a "hard/easy" effect in the perceptual domain replicating Baranski and Petrusic (1994, 1995, 1999) as well as Petrusic and Baranski (1997). Their findings support the position that the degree of under/over confidence is due to task difficulty and not illusory effects as suggested by Olsson and Winman (1996). Oakley (1998) also found a difficulty effect in the perceptual domain on four perceptual tasks for two age groups. Research showed that, on average, all participants showed decreased over-confidence, improved calibration (i.e. lower CI values) and improved resolution (higher resolution values) on easy items compared to hard items. The "hard/easy" effect has also been noted in the visual recognition memory domain by Oakley-McKeen and Petrusic (1995). Based on their research, Oakley-McKeen and Petrusic reported that difficulty in the task was associated with the "subjective ease of verbal codability" (Oakley, 1998, p. 43). These results were further replicated by Oakley (1998) in the general knowledge domain with the 'cities in Canada' task. Here, as with the perceptual tasks, easy items were associated with higher mean confidence, higher accuracy, under-confidence (or lower over-confidence), better

calibration and improved resolution compared to hard items, for young adults and seniors. Consequently, the results of research on difficulty effects identify the importance of varying the level of difficulty when conducting calibration research.

Because the literature on metacognition and AD seems to uniformly suggest that perceptual abilities remain intact longer than memory abilities (Raskind & Peskind, 1992; Mayeux & Schofield, 1994; Moss & Albert, 1992), the importance of including both a perceptual and a memory task is evident in order to determine whether or not domain specificity plays a role in the detection of early cognitive loss. Results from Oakley (1998) show no clear domain specific group differences between normal seniors and non-criteria seniors on individual perceptual tasks (i.e., Inspection Time) or memory tasks (i.e., general knowledge of cities in Canada); however, given the small sample size associated with the group of non-criteria seniors, results are not conclusive.

The controversy surrounding domain specificity as well as the suggestion that perceptual abilities remain intact longer than semantic memory supports the inclusion of both a memory task (general knowledge of cities in Canada) and a perceptual task (IT) in the current thesis. Also, it is clearly important to measure difficulty effects in order to ensure that differences observed are due to cognitive differences and not to task difficulty.

***Theories of confidence***

Three theories of confidence are relevant to the current thesis. Gigerenzer *et al.* (1991) proposed the *Probabilistic Mental Model* to account for over-confidence in the general-knowledge domain. This theory explains that over-confidence is a direct consequence of the experimental methodology and not of the individual. That is, if a

representative stimulus set (for example all states in the United States of America) is used then individuals should be well calibrated. The theory posits that when people are faced with a two-alternative-forced-choice (2-AFC) decision, they first try to construct a local mental model to enable them to solve the task through direct memory retrieval or rudimentary logic operations (e.g., exclusion). Failing this, the individual then constructs a probabilistic mental model, which employs probabilistic information from the natural environment to make the decision. For example, if it is not known which of two cities has the larger population, other information that is known about each city will be used in order to infer the information required to make the decision. For instance, if the task is to decide which city has the larger population, Oshawa, Ontario or Chicoutimi, Quebec, and the difference is not known, one may use the knowledge of sports teams (e.g., whether the city has national league or junior league sports), personal familiarity with the city or any other clue as to the difference between the sizes of the two cities. Confidence, according to this model, is based on the validity of this probabilistic cue. If the sample set is non-representative (biased) then the validity of the probabilistic cue is no longer accurate and over-confidence will be observed. Good calibration (i.e. a low CI value) is expected when the cue validity corresponds to its actual occurrence in the natural environment (i.e., when data are representative of an individual's natural environment then the individual is expected to be a good judge for the reliability of their knowledge).

The probabilistic mental model further suggests that if an individual is unable to invoke appropriate references then they will guess and indicate a confidence level of 50%. The resulting over-confidence and hard/easy effects are, according to this theory, functions of stimulus difficulty resulting from the sampling method (i.e., selected sample

vs representative sample). Given a representative sample, this theory predicts that people will be well calibrated and will show neither over- nor under-confidence. These hypotheses, however, have not been supported by the literature. In 1992, Griffin and Tversky found mis-calibration despite the use of a complete stimulus set of pairs of American states and more recently, Oakley (1998) observed over-confidence despite the use of a random selection of items from an entire set of cities in Canada. She noted that confidence was determined by stimulus difficulty as well as individual differences in self-confidence for young adults.

The probabilistic mental model was developed to explain the general knowledge domain only, and was not designed to account for results obtained in the perceptual domain where under-confidence is often reported (Keren, 1988; Bjorkman *et al.*, 1993; Baranski & Petrusic, 1994, 1995, 1999; Olsson & Winman, 1996; Petrusic & Baranski, 1997; Oakley, 1998). As a complement to the probabilistic mental model, Juslin and Olsson (1997) offered the *Sensory Sampling Model* as an attempt to explain subjective probability assessment in the sensory (perceptual) domain[2]. Both this model and the probabilistic mental model are based on the belief that perception and cognition involve different processes and, therefore, different origins of uncertainty (Juslin & Olsson, 1997). These are the Brunswikian and Thurstonian models. The PMM is based on the Brunswikian model which states that uncertainty arises from limited knowledge about the world gathered as partial pieces of information, or cuing, from the environment. The SSM, on the other hand, is based on the Thurstonian model which suggests that uncertainty is due to variability in the sensory system. In the Sensory Sampling Model, confidence is based on the proportion of sensations that are consistent with the decision

(e.g., 9 out of 10 sensations that are consistent with the decision results in 90% confidence). Juslin and Olsson predict that under-confidence is more likely to be observed in the perceptual domain due to the increased sensitivity of the decision over confidence. However, research has found over-confidence on difficult perceptual decisions (e.g., Baranski & Petrusic, 1994, 1995, 1999; Petrusic & Baranski, 1997, Oakley, 1998). This discrepancy may be due to the disclaimer made by Juslin and Olsson that under-confidence can only be predicted by this model when the stimulus duration is under the control of the subject, when the decision is not based upon unsuccessful inferences (errors), and when there are no biases in response or perception. In these circumstances, over-confidence will be observed. It would seem that these situations, though, are unnecessarily constraining and not the rule; thus, over-confidence should be observed more regularly than under-confidence.

Results from the Inspection Time perceptual task reported in Oakley (1998) do not support this theory. She found that over-confidence was observed for hard items and under-confidence for easy items in both positional orders, and at each SOA, for each of three groups. A discrepancy remains, therefore, about the prediction of performance by AD patients. The sensory sampling model would predict over-confidence in the Inspection Time task because the stimulus duration is experimenter-controlled. However, based on Oakley's results, the level of under/over confidence should vary with the difficulty of the stimulus and/or the perceived difficulty due to SOA. Such results, if obtained, would support the Doubt Scaling Model put forth by Baranski (1991).

The *Doubt Scaling Model* (Baranski, 1991) is an extension of Petrusic's (1992) evidence accrual based '*Slow and Fast Guessing*' model (Petrusic & Baranski, 1989a, b;

Baranski, 1991). This model suggests that the choice for one alternative over the other in a 2-AFC decision depends on the difficulty of the judgement and decisional criteria. This theory can account for differences in the rate at which evidence is accumulated and allows for flexibility in decisional criteria by the individual. This theory suggests that easy judgements will result in higher levels of confidence and higher accuracy as well as quicker reaction times (RTs). Conversely, hard judgements will result in lower confidence and accuracy as well as longer RTs. This theory is unique in that it accounts for the 'uncertainty' variable. If an individual is uncertain or indifferent then a guess will be made which will result in lower accuracy and longer RTs for difficult comparisons where uncertainty is higher. To summarize, "confidence is obtained by scaling the amount of inconclusive information (doubt) at the completion of the judgement" (Baranski, 1991). Over-confidence, then, increases as the difficulty of the judgement increases and under-confidence can occur when the decision is easy. Thus, the doubt-scaling model can account for both under- and over-confidence judgements and Baranski and Petrusic (1994), therefore, propose that memory and perceptual tasks invoke similar processes. Although this model does not make a priori predictions, it could accommodate increased over-confidence in AD patients due to poorer performance overall.

***Population of Cities in Canada and Inspection Time Tasks***

As part of a larger study being conducted by Dr. Oakley at the Elisabeth Bruyere Hospital, Ottawa, participants received each of six experimental tasks randomized using a Latin-square design. Because the research literature indicates deterioration in general knowledge in early stages of AD and perceptual abilities appear to remain intact longer, it is crucial to include both a perception task and a general knowledge task in the evaluation

of metacognitive abilities. Consequently, two of these experiments are being investigated in the current research: one general knowledge task (Populations of Cities in Canada) and one perceptual task (Inspection Time). The general knowledge task (i.e. Cities in Canada) was chosen from the battery of tests used by Oakley (1998) because literature on AD has reported very early impairment of semantic memory which is that component of long-term memory containing knowledge of objects, facts and concepts as well as words and their meanings (Hodges & Patterson, 1995; Norton et al., 1997). The Inspection Time task was selected because this task yielded clear age effects as well as sensitivity to early cognitive loss in older adults.

***Perceptual Task : Inspection Time.*** The examination of inspection time was selected as it provides a speeded stimulus presentation at various stimulus durations independent of time required for decision making. The task utilizes masked perceptual stimuli (line lengths) which vary in the length of time between the presentation of the stimulus pair and of the mask, known in the research literature as stimulus-onset-asynchrony (SOA). The presentation time of the stimulus is presumed to be limited to "one inspection of the sensory input". The term "inspection time" (IT), originally proposed by Vickers, Nettleback and Willson (1972) is believed to provide an index of the "speed of the brain's immediate reaction to sensory input -- in the absence of any requirement for thought" (Brand & Deary, 1982, p. 300). In particular, this experiment extends from research conducted on age differences in confidence by Oakley (1988). Oakley found that differences were observed between normal seniors and seniors who failed to meet a priori criteria (eg. mean Choice Reaction Time (CRT) < 1 second) suggesting that this task may be sensitive to early cognitive loss.

Consistent with difficulty effects observed by Baranski and Petrusic (1994, 1995, 1999) and Petrusic & Baranski (1997) with young adults, Oakley found that both groups of seniors yielded lower accuracy, lower confidence, and greater over-confidence for the Hard line length comparisons (95, 100 pixels) compared to the Easy comparisons (75, 100 pixels). Over-confidence was also observed for both groups at both levels of difficulty (at each level of SOA) challenging the basic assumption of the Sensory Sampling Model that under-confidence would always be observed in the perceptual domain. Results for calibration and resolution were in the expected direction; that is,

Calibration worsened (i.e. CI values increased) as did normalized resolution, $\eta^2$, (i.e. $\eta^2$ values decreased) as difficulty of the comparison increased.

***Pre-Measures***

The experimental tasks were preceded by a number of pre-measures. A choice response time (CRT) task was included to detect whether or not the participants are able to follow instructions and make a basic two-alternative-forced-choice (2-AFC) decision. Also included as pre-measures were a Stroop colour naming task, the Ishihara test for colour vision, a visual acuity measure, demographic data (e.g., age, gender, education, handedness) and neuropsycholological information were obtained from medical files.

*Demographic data*[3] and any other personal information required for the study was obtained from medical files following the completion by the participant or his/her legal representative of two "Consent to the Disclosure of Personal Information" forms, (see *Appendices A, B*), one allowing access to medical files and the other explaining the study and the rights of the participant (i.e., the right to withdraw from the study).

*Two alternative Choice Response Time (CRT).* The CRT task requires participants to indicate on a response panel whether a small vertical line appears to the left or to the right of a central fixation point on the computer screen, over 64 trials divided into 4 sets. (see complete instructions, *Appendix C*).

*Stroop Colour Naming.* The Stroop task (Stroop, 1935), also computerized used three stimuli (the word "red", the word "green" or three stars "***") presented randomly. The stimuli will be presented in either the colour red or the colour green requiring participants to indicate on the response panel whether the colour of the print of the word or stars was red or green in colour using buttons marked "R" for red or "G" for green.

Trials on which the colour and the word match (e.g., the word "red" is presented and the colour is also red) are considered congruent trials, whereas incongruent trials are those trials on which the colour and the word do not match (e.g., the word "green" is presented in the colour red). Participants will be instructed to make their responses as quickly and as accurately as possible with the total task time being approximately 5-10 minutes (see instructions, *Appendix D*). The Ishihara test of colour blindness was administered to ensure that all participants are able to distinguish between the colours red and green (Ishihara, 1968).

*Snellen Test of Visual Acuity.* As all of the tasks involved in this project are visual in nature, the Snellen test of visual acuity, used at the recommended distance of 20 feet, was employed to ensure that visual acuity impairment was not a factor in the performance of participants. Participants were requested to wear any corrective eyewear for the experiment.

***Calibration and Alzheimers Disease: Hypotheses***

In an attempt to identify differences between Non-AD seniors and those diagnosed with AD, data on seniors from Oakley (1998) will be compared to a group of seniors who have been identified as having early stages of either probable or possible AD on each of a memory task (Cities in Canada) and a perceptual task (Inspection Time).

As difficulty effects have been noted regularly in both general knowledge and perceptual domains, difficulty levels will be manipulated for the Cities in Canada and the Inspection Time tasks. In accord with the Doubt Scaling Model, higher levels of confidence, higher accuracy and quicker reaction times are expected for easy items compared to hard items.

In order to directly test the possibility that the CI may be useful in detecting cognitive loss due to AD, as suggested by Oakley's 1998 findings, values on the CI will be reviewed for seniors not diagnosed with AD and seniors who have been diagnosed with AD on both a general knowledge task (Cities in Canada) and a perceptual task (Inspection Time). Although results from Oakley (1998) did not show clear domain specific group differences between seniors who met a-priori criteria and seniors who did not on memory and perceptual tasks, results were not conclusive given the small sample size associated with the non-criteria group. The literature on metacognition and AD, however, suggests that perceptual abilities remain intact longer for AD seniors than their memory abilities (Raskind & Peskind, 1992; Mayeux & Schofield, 1994; Moss & Albert, 1992) and that AD seniors differ from normal seniors in their inability to remember (Masur *et al.*, 1989). Given this, it could be hypothesized that differences will be seen between groups on the memory task but not the perceptual task given the suggested difference in onset of symptoms. However, given Oakley's findings, domain specificity may not be observed. Performance abilities are expected to differ between groups, however, with AD seniors being less accurate than non-AD seniors on the memory task. In agreement with the Doubt Scaling Model put forth by Baranski (1991) and the Oakley's 1998 findings, over-confidence is expected on hard items, and under-confidence on easy items on the perceptual task for both groups.

Based on findings from Oakley's 1998 study as well as findings from the literature, the hypotheses of the current thesis are as follows:

1) A "difficulty effect" is expected on both the memory and perception task (with higher accuracy, higher mean confidence, less over-confidence, better calibration (lower CI values) and better resolution (higher resolution values) on easy items compared to hard.
2) Based on the Doubt Scaling Model, over-confidence is expected on hard items and under-confidence on easy items on the perceptual task for both groups.
3) As memory is thought to be affected in those with AD more than in normal seniors and as memory abilities are thought to be affected before perceptual abilities, group differences are expected on the memory task in that AD seniors are expected to be less accurate, more poorly calibrated (higher CI values) and more poorly resolved (lower values for resolution) than normal seniors.
4) Given Oakley's (1998) findings, group differences in both performance and calibration could also be expected for the perceptual task.

## METHOD

*Participants.* Participants were 13 older adults (age 55 +) suffering from either "Possible" or "Probable" Alzheimer's Disease. Generally, individuals with this diagnosis show mild to moderate stages of cognitive decline with memory problems that are apparent to others and that interfere with the ability to complete complex task. Based on a power analysis performed for the larger study from which this experiment is drawn, the sample size sought to achieve a power level of .8 was 50 participants; however this was not achievable due to limitations of recruiting from a clinical population. Participants were drawn from the patient pool in the Memory Disorder Clinic at the Elisabeth-Bruyere Centre (SCO hospital). Patients with mild to moderate cognitive impairment were sent a

letter from the Director of the Memory Disorder Clinic (see *Appendix E*) explaining the study and subsequently contacted by telephone for recruitment. Clients were offered remuneration for out-of-pocket expenses (eg. parking, light lunch). Exclusion criteria included clinical depression, severe visual impairment (worse than 20/40 visual acuity), severe arthritis in the hands, or a recent history of violence. The participants in the Non-Alzheimer's Seniors (Non-ADS) group were the 52 seniors employed in Oakley (1998).

*General Knowledge Task: Population of Canadian Cities*

*Stimuli.* The stimuli were presented in the manner described in Oakley (1998). From the Canadian Almanac (1990), fifty names of Canadian cities were used to generate a large data set of pairs of city names. Vancouver, Toronto, and Montreal were not used as they are well-known metropolitan areas. The city of Ottawa was also excluded as this was the city in which data collection took place, which would have made comparisons, presumably, too easy. Each city name was paired with each of the other city names except itself, resulting in 1225 pairs of cities using the combinatorial coefficient (50 x 49/2). The ratio of the larger city population to the smaller for each pair was computed and used to determine the level of stimulus difficulty. Pairs of city names, rank ordered based on ratio, was divided into three sets with three corresponding levels of difficulty. The third with the smallest ratios (< 1.5) was considered the Hard set, the Intermediate set consisted of the third in the centre of the rank ordered list (i.e., ratios between 1.5 and 2.3) and the final third was considered the Easy set (i.e., ratios larger than 2.3). The Intermediate level of difficulty was used for practice trials only, and the Hard and Easy difficulty levels were used for the experimental trials.

For the practice trials, eight pairs of cities were randomly selected from the intermediate set. Sixteen pairs from the Hard and from the Easy sets were randomly selected for each experimental block. Half of the pairs of each set were presented with the instruction "larger" for which the subject decided which city name they thought corresponded to the instruction, that on the left or that on the right (see *Figure 1*).

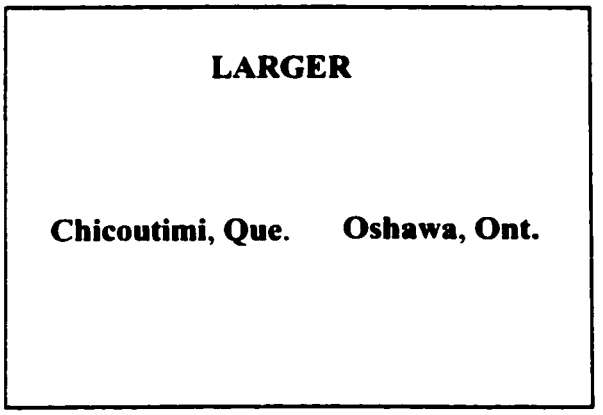


*Figure 1.* Stimuli - Cities in Canada Experiment

The remaining stimulus pairs were presented with the instruction "smaller" where the selection was made as to which city, the one on the left or the one on the right, was the smaller of the two. In order to ensure that the larger city was not always on the left (because the data were generated using a factorial design), the positional-order of half of each set was reversed (four per instruction). This resulted in 32 pairs of cities in each of two blocks of trials (2 levels of difficulty x 2 instructions x 8 pairs) for a total of 64 stimulus pairs.

*Procedure.* In the 2-AFC general knowledge task, participants were asked to make the decision as to which of a pair of cities had either the larger or the smaller population. On each trial, either the instruction "LARGER" or "SMALLER" was

presented at the top centre of the screen followed by the simultaneous presentment of a pair of cities. The participants indicated their responses from memory by depressing a key marked "L" if they thought the city on the left corresponded to the instruction, or "R" if they thought the city on the right corresponded to the instruction. Participants were then asked to indicate their level of confidence about the decision they had just made by depressing a button on the response panel that corresponded to their level of confidence (0 for a mistake, 50 for a guess, 60 to 90 indicating increasing confidence and 100 indicating certainty). Participants were asked to respond as quickly and as accurately as possible (see complete instructions: Appendix F). Two blocks of trials were given with a short rest in-between. The entire task took approximately 20-25 minutes to complete.

*Perceptual Task: Inspection Time*

*Stimuli.* A computer presented stimuli consisting of pairs of vertical line lengths connected at the top by a bar, and the instruction of either "longer" or "shorter" appeared at the top centre of the screen (see *Figure 2*). The brief presentation of the stimulus was then masked with a thicker two-legged T-bar with equal vertical line lengths.

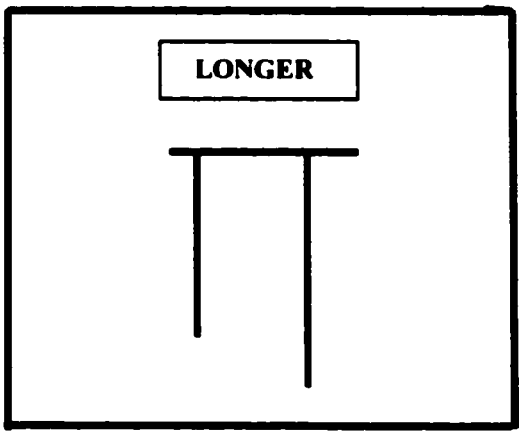


*Figure 2.* Stimulus display for the inspection time task.

The discrimination difficulty levels were manipulated by presenting lines that were either very close in length (5% difference -- Hard condition), or very different in length (25% difference -- Easy condition). Difficulty was also manipulated by the duration of the SOA, which varied based on multiples of the raster-scan rate of the computer monitor (2, 4, 6, 8, 10, 12). The raster scan for the Samtron monitor is 16.686 ms; therefore SOAs of 33.372, 66.744, 100.116, 133.488, 166.860 and 200.232 ms were used with SOAs of 2 and 4 considered Short, 6 and 8 Moderate, and 10 and 12 Long. At the end of each of two blocks of trials, the participant was asked to indicate whether or not they saw any "movement" in one of the lines that they used to help them make a decision. Participants were asked to indicate the percentage of trials on which they observed movement (from 0 -100%) using the computer keyboard. Next, individuals who reported seeing movement were asked to indicate the magnitude of that movement on a scale from 1-9. The phenomenon of reporting motion in one of the lines was identified in the literature as a potential cognitive strategy by some individuals for decision-making. This measure is included in the current thesis in order to examine sources of individual differences in calibration and to replicate the procedure used by Oakley (1998).

Participants first received a practice block consisting of 48 practice trials. The two experimental blocks followed which consisted of 144 trials each (2 instructions x 2 pairs x 2 position orders x 6 SOAs x 3 replicates). Participants were given a self-paced rest-break between blocks. In further replication of the study by Oakley (1998), participants were instructed to be as accurate as possible in making their decision (see *Appendix G*).

## RESULTS AND DISCUSSION

The findings are presented in two main sections; the first examines the preliminary CRT and Stroop measures with a view to ensuring that, indeed, the performance of AD participants was in fact significantly impaired. The second provides detailed analyses of the cognitive based cities population discrimination task and the perceptually based line length, inspection time task, in that order. Within each of the latter, main tasks, the global measures of discriminative accuracy (percent correct) and mean confidence (percent confidence) are first examined, followed by each the various calibration analyses indices. Finally, Response Time is examined with a view toward specifying cognitive loss in terms of processing speed for the AD group relative to the Non-Ad Group.

Throughout, levels of significance for the repeated measures factors were determined using the Huynh-Feldt epsilon adjusted degrees of freedom, although the degrees of freedom indicated in the text are defined by the design.

### *Preliminary Tasks*

*Choice Reaction Time.* As the plots in Figure 3 show, CRTs of the 13 AD seniors participating in this task are substantially slower than those of their 52 Non-AD counterparts; mean CRT for the AD participants is 942.9 ms compared to 497.6 ms for the Non-AD seniors. The differences between mean CRTs for the two groups were highly reliable ($F(1, 63)=8.16$, $MS(Error)=1{,}011{,}291.76$, $p<0.0058$. However, no other main effects nor interactions attained conventional statistical significance. Thus, these data confirm significant cognitive slowing on an elementary choice reaction time task with the AD participants.

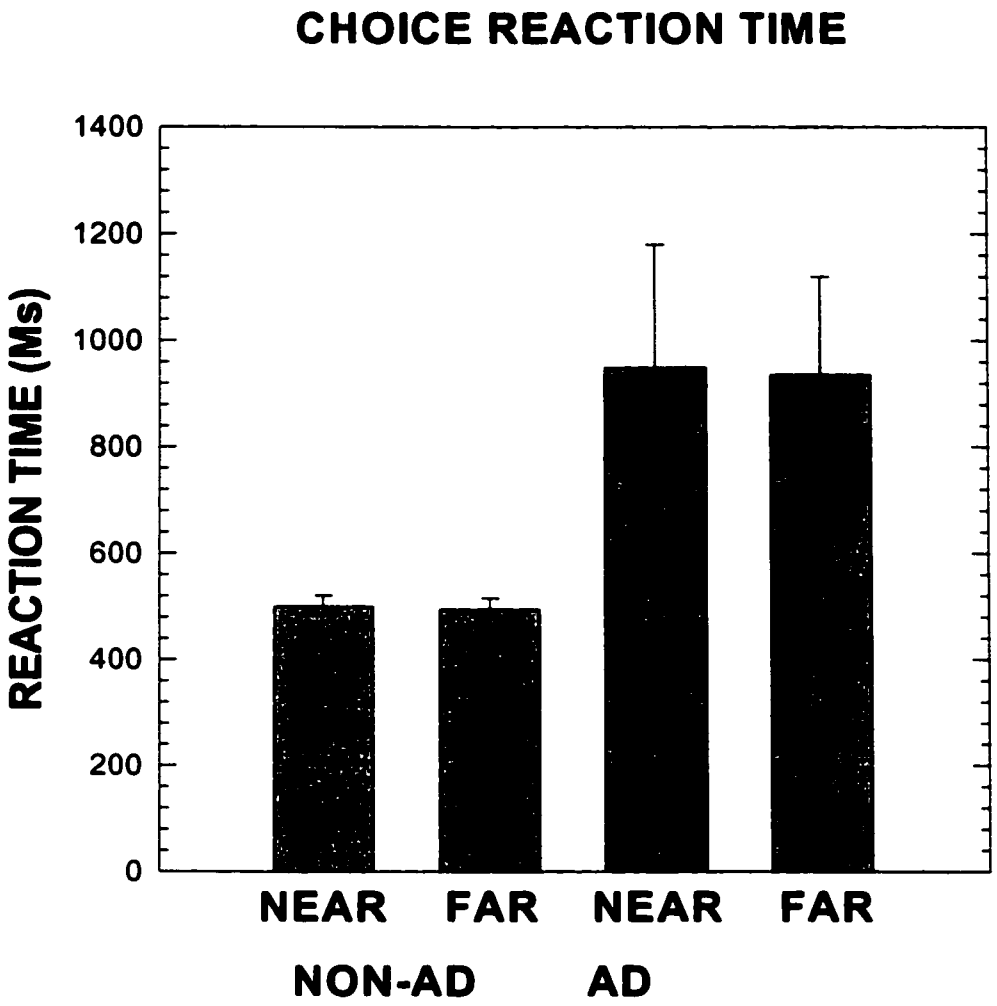


*Figure 3.* Choice reaction times for the Non-AD seniors and Alzheimers Disease participants with the near and far stimuli.

*Stroop Colour Naming.* As in the CRT task, the 15 AD seniors participating in this task showed significantly slower overall mean colour naming times (see *Figure 4*); AD participants took 1006.1 ms to respond compared to 755 ms for the 53 participating Non-AD seniors. Relative to the neutral condition, AD participants were slowed by 272.99 ms with the incongruent letter string compared to 206.3 for the Seniors. Interestingly, Seniors show virtually no facilitation (3.3 ms) as is typical of college student data (see MacLeod, 1992 for a review); the AD participants showed 80.8 ms of facilitation. However, the interaction between group and condition fails to attain conventional statistical significance ($F(2, 132)=2.48$, $MS(Error)=24265.29$, $p<0.0927$). Thus, the Stroop colour naming data converge nicely with the CRT task findings of general and pervasive cognitive slowing. As well, the heightened failure to inhibit the competing semantic information for the AD participants is in the predicted direction. Although this effect failed to meet significance, the findings are in line with earlier work documenting increasing inefficiency in inhibitory control with increasing age (Hasher & Zacks, 1998; McDowd & Filion, 1995).

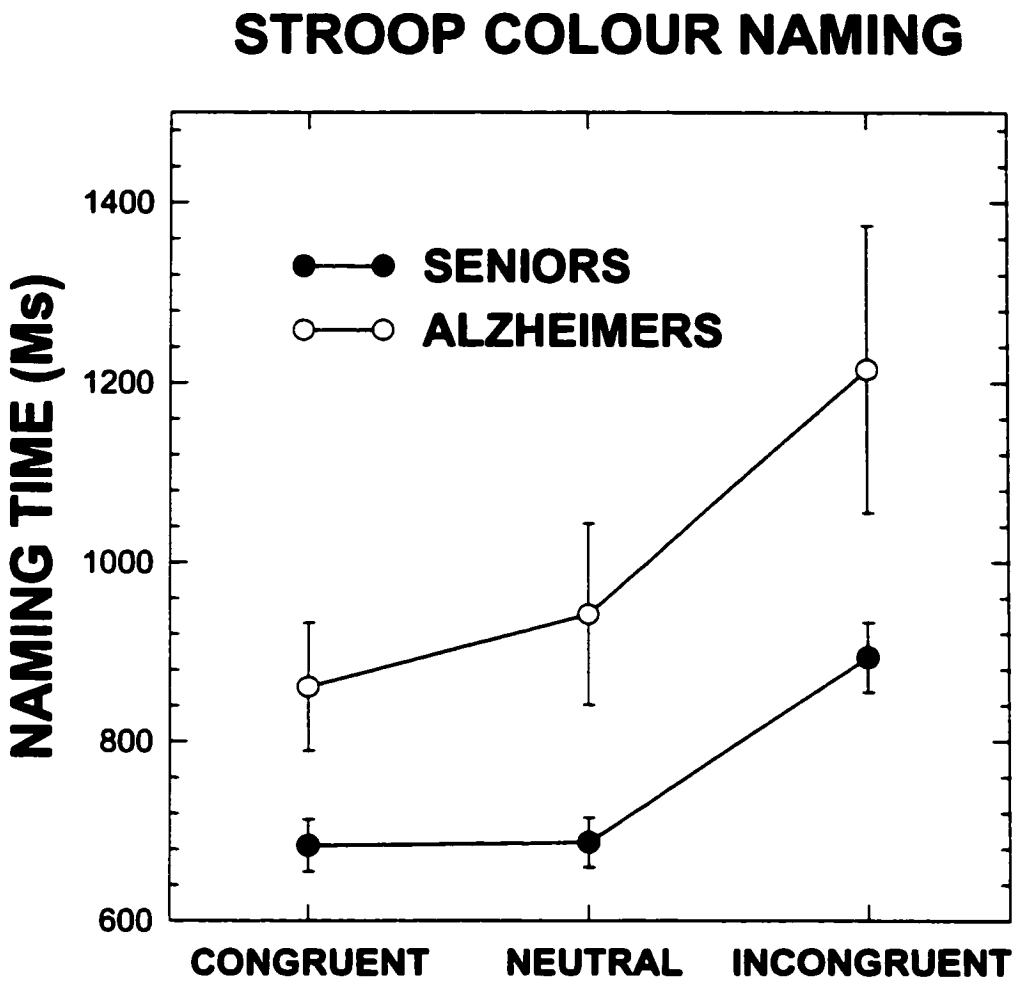


*Figure 4*. Mean colour naming RTs for Non-AD seniors and Alzheimers participants for congruent, neutral, and incongruent conditions.

***General Knowledge Task: Population of Canadian Cities***

*Percent Correct and Percent Confidence.* The mean percentage correct for the 53 Non-AD Seniors participating in this task was 72.5; it was 60.8% for the 13 participating AD seniors and this difference is statistically reliable ($F(1, 60)=27.44$, $MS(Error)=101.01$, $p < 0.00001$). AD participants were less confident overall (69.6%) than Non-AD participants (78.3%) and this difference in overall mean confidence misses conventional significance ($F(1, 60)=3.64$, $MS(Error)=425.86$, $p<0.061$).

The main effect of difficulty was, as expected, highly reliable ($F(1, 60)=220.77$, $MS(Error)=60.58$, $p<0.00001$); for the difficult comparisons mean overall accuracy was 55.6% and it was 84.5% for the easy comparisons. Correspondingly, confidence was lower for the difficult items (72.9%) than for the easy comparisons (80.1%) and this difference was also reliable ($F(1, 60)=33.45$, $MS(Error)=18.21$, $p<0.00001$). These main effects of difficulty measures were qualified by reliable interactions with the group factor for both the accuracy ($F(1, 60)=11.66$, $p<0.0011$) and the confidence ($F(1, 60)=9.71$, $p<0.0028$) measures. For the accuracy measure, the effects of difficulty were greatly amplified for the Non-AD participants (the difference is 31.4%) relative to the AD participants (here the difference is but 19.7%). Similarly, for the confidence measure, the differences between the hard and the easy comparisons was 8.4% for Non-AD participants but it was only 2.5% for AD participants.

*Calibration analyses.* Overall, both groups showed the typical over-confidence (6.4%) associated with general knowledge cognitive tasks ($F(1, 60)=10.51$, $MS(Error)=420.02$, $p<0019$), as is evident in the plots provided in Figure 5. As is also evident in Table 2, although the two groups did not differ in over-confidence for the

difficult comparisons they did for the easy ones. The Non-AD seniors were under-confident, as typically obtained for easy tasks, but the AD participants showed neither under nor over confidence. However, this interaction between difficulty and group failed to attain conventional statistical significance ($F(1, 60)=3.46$, $MS(Error)=50.99$, $p>0.068$).

As is evident in Figure 5 and Table 2, neither group is particularly well calibrated (i.e. low CI values), for either easy or hard comparisons. It is also clear that the two groups do not differ in calibration. However, there is a suggestion that the Non-AD seniors do show superior (albeit not that good either) normalized resolution (i.e. higher values for $\eta^2$). Indeed with the limited power available, the difference between the groups approaches significance for the normalized resolution measure, $\eta^2$, ($F(1, 60)=3.30$, $MS(Error)=0.015$, $p<0.074$).

| Group | Conf | p(c) | O/U | CI | Res | $\eta^2$ |
|---|---|---|---|---|---|---|
| | | | EASY | | | |
| **Non-ADS** | 82.50 | 88.13 | -5.63 | 0.036 | 0.009 | 0.131 |
| **ADS** | 70.88 | 70.66 | 0.22 | 0.035 | 0.008 | 0.046 |
| | | | HARD | | | |
| **Non-ADS** | 74.12 | 56.76 | 17.36 | 0.079 | 0.012 | 0.051 |
| **ADS** | 68.37 | 51.01 | 17.36 | 0.077 | 0.008 | 0.038 |

*Table 2*. Mean Confidence (conf), mean Proportion Correct (p(c)), Under/Over-Confidence (U/O), Calibration Index (CI), and Normalized Resolution ($\eta^2$), for Easy and Hard comparisons on the general knowledge of Cities in Canada task for Seniors who have not been diagnosed with AD (Snrs) and Seniors who have been diagnosed with either probable or possible AD (AD).

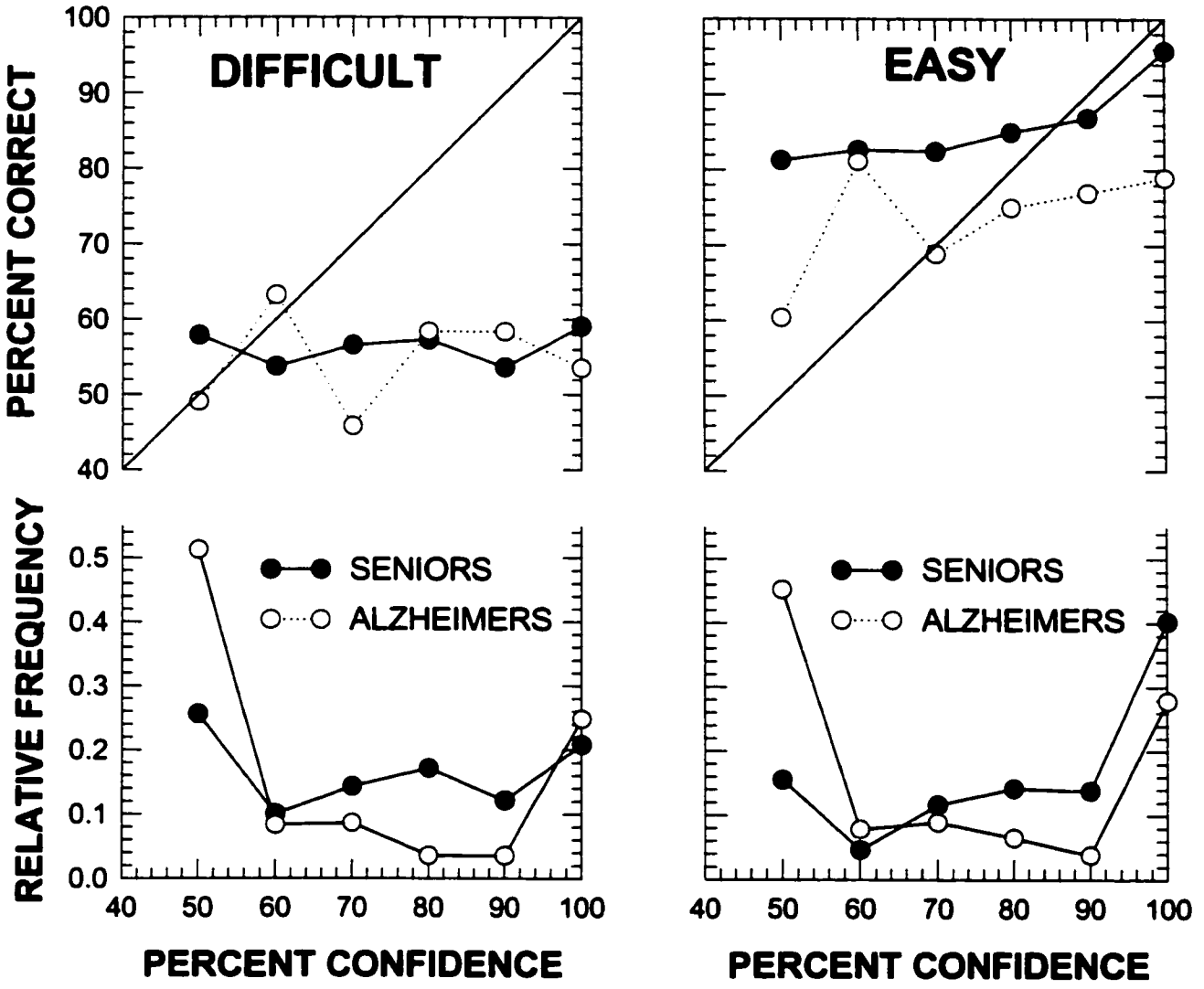


*Figure 5*. Calibration curves (above) and relative frequency curves (below) for two levels of comparative difficulty (hard and easy) on the Cities in Canada task for AD and Non-AD seniors groups.

*Response Times.* Response times were collapsed across both correct and error responses for all tasks. Figure 6 provides a view of the response time data for the AD and non-AD participants, with overall mean response times, plotted at each confidence category. Consistent with the cognitive slowing evident on the preliminary CRT and Stroop tasks, AD participants were significantly slower on the Cities-Populations task ($F(1,60) = 15.82$, MS(Error)=9,833,108.0, $p < .001$) than their non-AD Senior counterparts. Indeed, as the plots in Figure 6 show, at each level of confidence, the AD-participant's response times are clearly longer than those of the non-AD participants.

The pattern of response times in *Figure 6* reflects the classic finding, evident since the landmark work of Henmon (1911) and Johnson (1939), that RTs decrease as confidence increases (see also Baranski & Petrusic, 1994). It is also clear in the plots in Figure 6, that RTs are longer for the Hard pairs than for the Easy pairs ($F(1, 60)=60.37$, MS(Error)=466923.7, $p<0.00001$) for both the non-AD and AD seniors. This latter result is especially important in dramatically constraining the available quantitative theories of confidence and response time (cf., Baranski & Petrusic, 1994). The interaction between group and difficulty failed to meet significance, however.

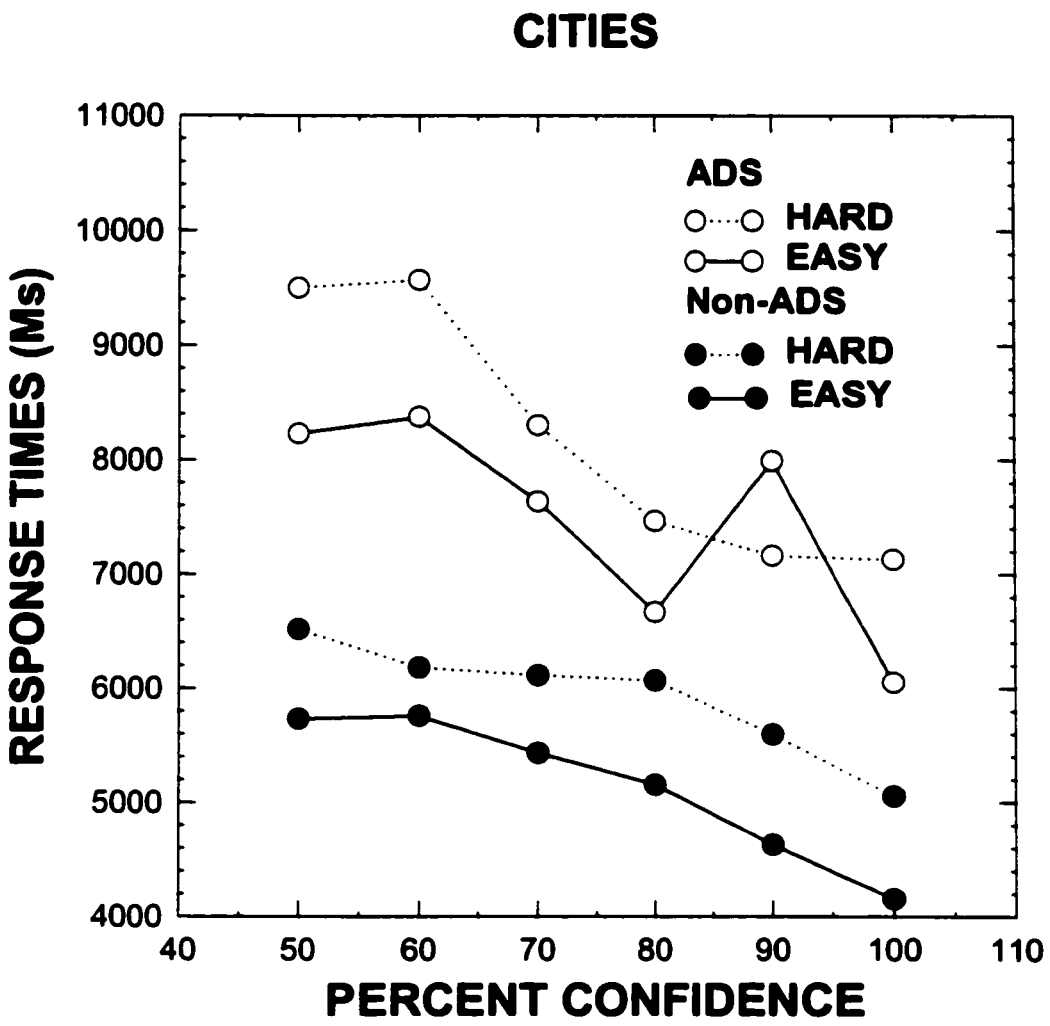


*Figure 6.* Response times by confidence category for Easy items (solid lines) and Hard items (dashed lines) on the Cities in Canada task for Seniors who have not been diagnosed with AD (Non-ADS) and Seniors who have been diagnosed with probable or possible AD (ADS).

***Perecptual Task: Inspection Time***

*Percent Correct and Percent Confidence.* The mean percent correct for the 49 Non-AD seniors completing the task was 71.2 (*Table 3*); it was 56.5% for the 9 AD participants, a difference that is statistically reliable ($F(1,57) = 18.39$, MS(Error) = 535.25, $p < 0.00001$). AD seniors were also less confident (63.5%) than Non-AD seniors (81.5%), a difference that is also statistically reliable ($F(1,57) = 11.46$, MS (Error) = 1293.19, $p < 0.001$).

As expected, the main effect of difficulty was highly reliable ($F(1,57) = 49.70$, MS(Error) = 100.28, $p < 0.00001$) with greater accuracy observed on easy items (75.1%) than on hard items (62.8%). Similarly, confidence was lower on hard items (76.3%) than on easy items (81.3%). This difference was also reliable ($F(1,57) = 20.46$, MS(Error) = 55.38, $p < 0.00001$).

This effect is also observed when broken down by SOA (*Table 4*) with greater accuracy seen on long SOAs (76.9%) than on medium SOAs (71.1%) and greater accuracy on medium SOAs than on short SOAs (58.8%); a reliable difference ($F(2,114) = 42.02$, MS(Error) = 75.07, $p < 0.00001$). This pattern is repeated when looking at confidence, with participants indicating greater confidence on long SOAs (81.8%) than medium SOAs (80.2%) and greater confidence indicated on medium SOAs than on short SOAs (74.3%). This difference is also statistically reliable ($F(2, 114) = 10.53$, MS(Error) = 49.90, $p < 0.001$).

The main effects of group and SOA are qualified by the interaction for accuracy ($F(2, 114)=7.46$, MS(Error) = 75.07, $p < .002$). Here, the effects of difficulty were greater among Non-AD seniors (where the difference between long and short SOAs was 20%) as

| Group | Conf | p( c ) | O/U | CI | Res | $\eta^2$ |
|---|---|---|---|---|---|---|
| | | | EASY | | | |
| **Non-ADS** | 84.031 | 77.7538 | 6.2764 | 0.0443 | 0.0151 | 0.1134 |
| **ADS** | 66.0253 | 60.366 | 5.6593 | 0.0594 | 0.0308 | 0.157 |
| | | | HARD | | | |
| **Non-ADS** | 79.0442 | 64.5951 | 14.449 | 0.0788 | 0.0195 | 0.1014 |
| **ADS** | 61.0602 | 52.6543 | 8.4059 | 0.0524 | 0.0255 | 0.1085 |

*Table 3.* Mean Confidence (conf), mean Proportion Correct (p(c)), Under/Over-Confidence (U/O), Calibration Index (CI), Resolution (Res) and Normalized Resolution ($\eta^2$), for Easy and Hard comparisons on the Inspection Time task for Seniors who have not been diagnosed with AD (Non-ADS) and Seniors who have been diagnosed with either probable or possible AD (ADS).

| | HARD | | | EASY | | |
|---|---|---|---|---|---|---|
| Group | SHORT | MEDIUM | LONG | SHORT | MEDIUM | LONG |
| Proportion Correct | | | | | | |
| **Non-ADS** | 55.6743 | 64.1579 | 73.9532 | 64.3771 | 82.7881 | 86.0962 |
| **ADS** | 49.0939 | 53.4162 | 55.4528 | 54.3452 | 62.8935 | 63.859 |
| Mean Confidence | | | | | | |
| **Non-ADS** | 75.3829 | 80.0043 | 81.7452 | 77.7569 | 86.3202 | 88.0159 |
| **ADS** | 60.4425 | 60.7897 | 61.9485 | 63.4226 | 66.7876 | 67.8657 |
| Under/Over Confidence | | | | | | |
| **Non-ADS** | 19.7086 | 15.8464 | 7.7921 | 13.3797 | 3.5312 | 1.9184 |
| **ADS** | 11.3486 | 7.3734 | 6.4957 | 9.0774 | 3.8941 | 4.0065 |
| Calibration Index | | | | | | |
| **Non-ADS** | 0.1038 | 0.0832 | 0.0495 | 0.0683 | 0.0341 | 0.0306 |
| **ADS** | 0.0866 | 0.0403 | 0.0304 | 0.0722 | 0.0587 | 0.0472 |
| Resolution | | | | | | |
| **Non-ADS** | 0.0169 | 0.0225 | 0.0192 | 0.0188 | 0.0146 | 0.0118 |
| **ADS** | 0.0402 | 0.0227 | 0.0136 | 0.041 | 0.0265 | 0.025 |
| Normalized Resolution | | | | | | |
| **Non-ADS** | 0.0712 | 0.1061 | 0.1268 | 0.0878 | 0.1331 | 0.1192 |
| **ADS** | 0.1756 | 0.093 | 0.0569 | 0.1824 | 0.1512 | 0.1375 |

*Table 4.* Mean Confidence (conf), mean Proportion Correct (p(c)), Under/Over-Confidence (U/O), Calibration Index (CI), Resolution (Res) and Normalized Resolution ($\eta^2$), for comparisons at short, medium and long SOAs on the Inspection Time task for Seniors who have not been diagnosed with AD (Non-ADS) and Seniors who have been diagnosed with either probable or possible AD (ADS).

compared with AD participants (difference of only 8%) (see *Figure 7*). This interaction was not found for confidence.

No interaction of group and difficulty was found for either accuracy or confidence. However, the difficulty effect was further qualified by the interaction of difficulty and SOA for both of these measures ($F(2, 114)=3.26$, MS(Error) = 58.63, $p < .04$ for accuracy and $F(2, 114) = 6.74$, MS(Error) = 8.97, $p < .004$ for confidence) with participants being most accurate (mean = 82.7) and most confident (mean = 84.9) on easy items with the longest SOAs (mean = 82.7) and least accurate (mean = 54.7) and least confident (mean = 73.1) on hard items with the shortest SOAs (mean = 54.7).

No interaction of group x SOA x difficulty was found for either confidence or accuracy.

*Calibration Analyses.* Overall, both groups showed over-confidence (9.9%) on this perceptual task ($F(1,57) = 12.12$, MS(Error) = 1142.12, $p < 0.001$).

Groups differed in that AD seniors were better resolved (mean = 0.028) than Non-AD participants (mean = 0.017). However, this difference failed to meet significance ($F(1,57) = 3.37$, MS(Error) = 0.0016, $p < 0.07$). No differences were observed between groups on calibration or over/under confidence however, as can been seen in *Figure 8*, a difficulty effect was observed with participants being less over-confident on easy items (mean = 6.2) than on hard items (mean = 13.5). This effect was statistically reliable ($F(1,57) = 10.6$, MS(Error) = 128.64, $p < 0.0019$).

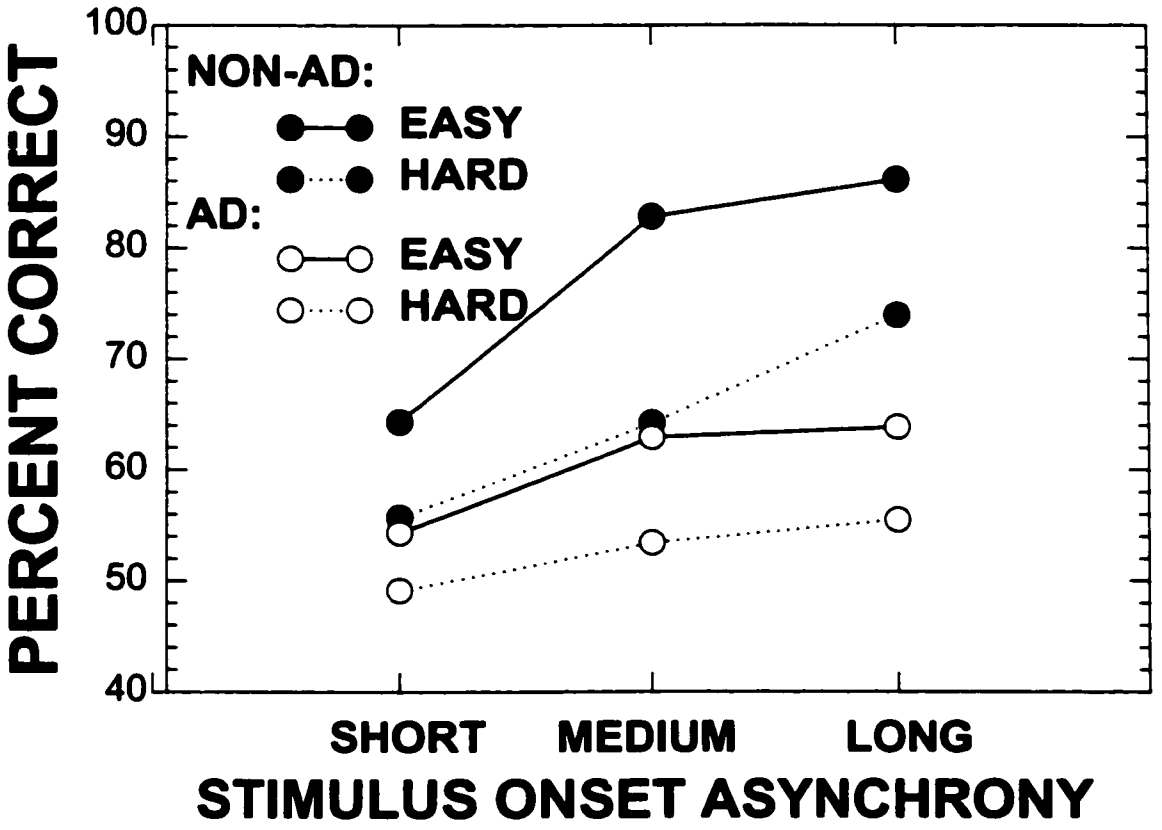


*Figure 7.* Accuracy (percent correct) at three levels of SOA (short, medium and long) at two levels of difficulty (hard and easy) for Non-AD seniors and AD seniors.

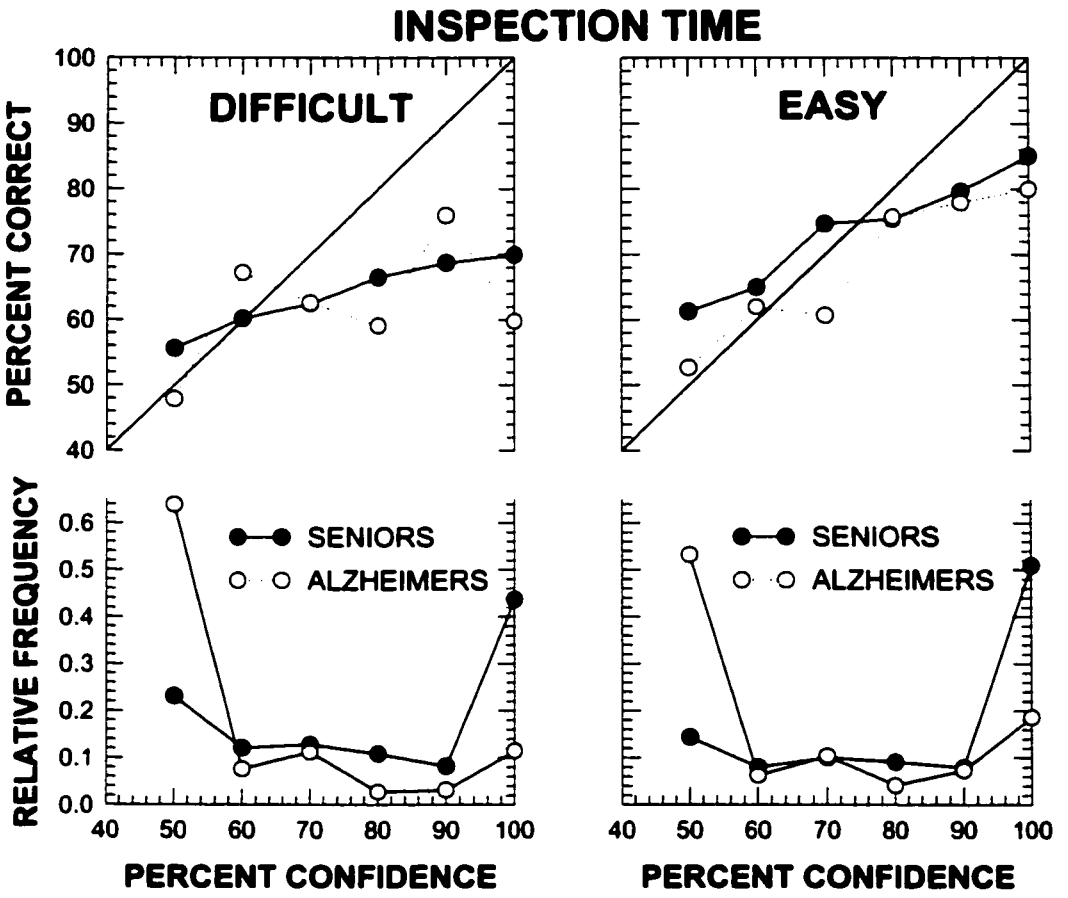


*Figure 8.* Calibration curves (above) and relative frequency curves (below) for two levels of comparative difficulty (hard and easy) on the Inspection Time task for AD and Non-AD seniors groups.

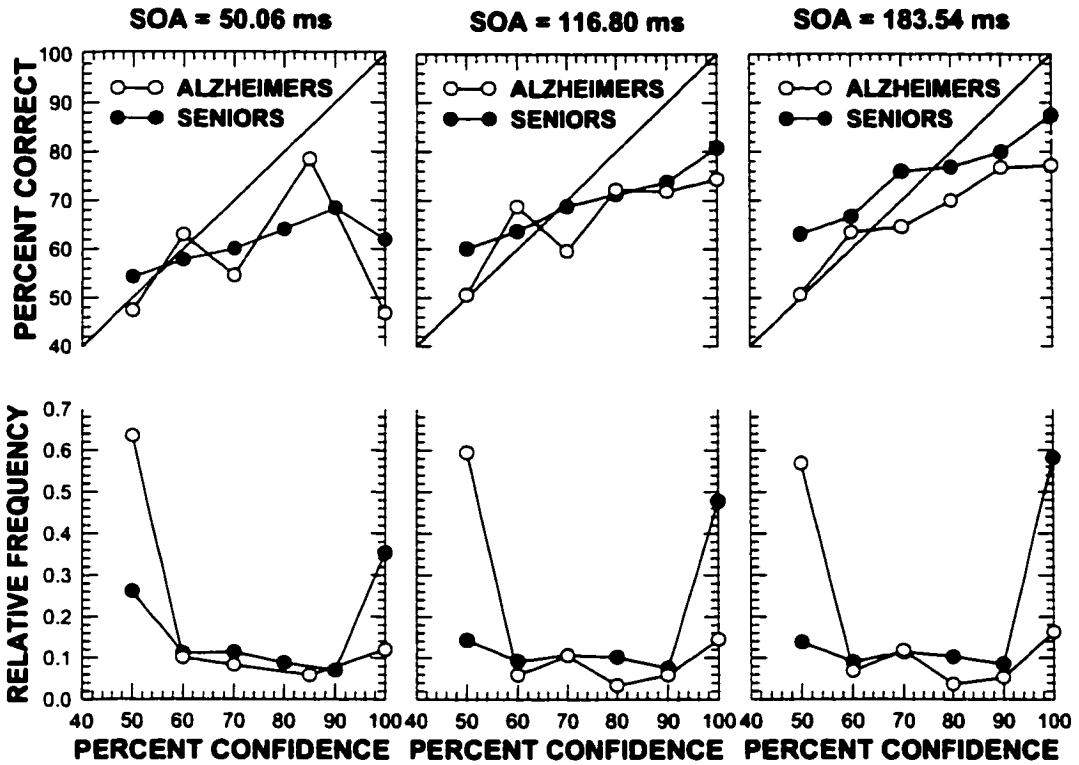


*Figure 9.* Calibration curves (above) and relative frequency curves (below) for three levels of SOA (short, medium and long) on the Inspection Time task for AD and Non-AD seniors groups.

As can be seen in *Figure 9*, this effect persists when breaking down the task by SOA ($F(2,57) = 10.13$, MS(Error) = 109.26, $p < 0.0004$) with participants being less over-confident on long SOAs (4.9%) than on medium SOAs (9.1%) and less over-confident still on medium SOAs than on short (15.6%). The effect of SOA continues to be observed with participants being better calibrated (i.e., lower CI values) ($F(2,57) = 9.65$, MS(Error) = 0.003, $p < 0.0005$) and resolved (i.e. higher resolution values) ($F(2,57) = 6.83$, MS(Error) = 0.0003, $p < 0.002$) on long SOAs (mean calibration = 0.004; mean resolution = 0.0161) than on medium SOAs (mean calibration = 0.06; mean resolution = 0.0195) and still better still on both indices on medium SOAs than on small SOAs (mean calibration = 0.09; mean resolution = 0.0213). This effect was not seen once resolution scores were normalized, however.

When exploring the interaction of group and SOA, there are obvious differences, with Non-AD seniors being better resolved (for normalized scores) as SOAs become longer (mean = 0.079 for short SOAs, 0.119 for medium SOAs and 0.123 for long SOAs) but with AD participants showing the reverse patters (mean = 0.179 for short SOAs, 0.122 for medium SOAs and 0.097 for long SOAs). This interaction was reliable ($F(2,114) = 4.30$, MS(Error) = 0.025, $p < 0.02$). This suggests that AD seniors use the confidence categories to indicate whether they are correct or not more effectively than their Non-AD counterparts. This is particularly the case when the task is difficult, as can be seen from Figure 10.

The lack of a group effect for calibration and the presence of a group effect for resolution were qualified by a group by difficulty interaction for both calibration ($F(1,57) = 6.10$, MS(Error) = 0.003, $p < 0.02$) and resolution ($F(1.57) = 5.04$, MS(Error) = 0.0002,

$p < 0.03$) but not for normalized resolution. Table 3 and Figure 5 show Non-AD participants as being better calibrated on easy items (0.044) than on hard items (0.079), whereas AD seniors are essentially equally poorly calibrated on both easy (0.059) and hard (0.052) items. Alternately, this table and figure also show AD participants as better resolved on easy items (0.0308) than on hard items (0.0255) and Non-AD seniors being equally poorly resolved on both easy (0.0151) and hard (0.0195). These findings are interesting in that AD participants appear to have an awareness that they are unable to perform this task at the most difficult levels whereas their Non-AD counterparts remain poorly resolved despite their inabilities at this degree of difficulty.

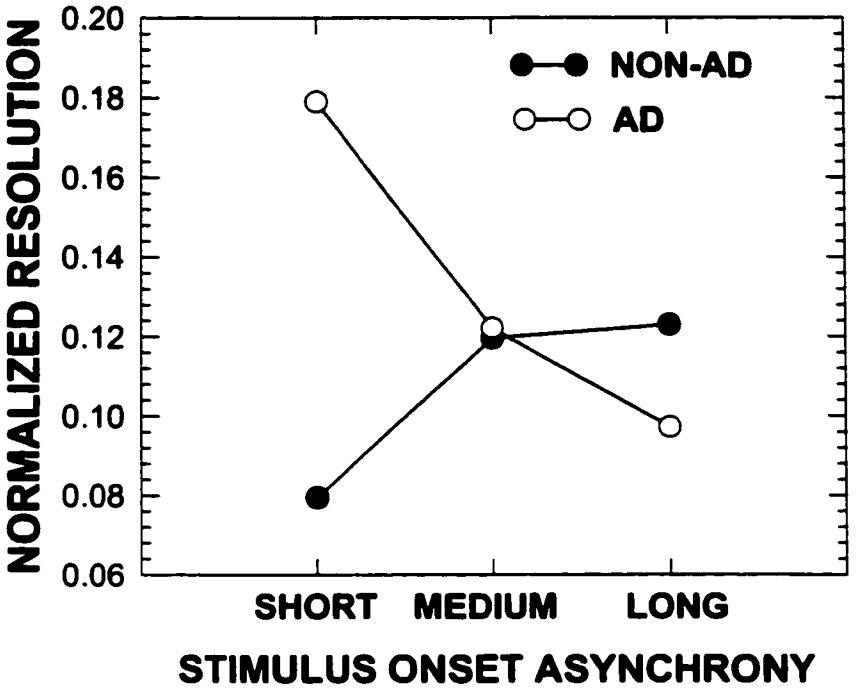


*Figure 10.* Normalized Resolution of Non-AD and AD participants at three lengths of Stimulus Onset Asynchrony (SOA) for the Inspection Time task.

No interaction effects were observed for difficulty by SOA on any of the calibration indices, nor were there any effects of SOA by group by difficulty.

*Reaction Time.* As was evident in the CRT and Stroop pre-measures, cognitive slowing in AD Seniors was also observed in participant response times. As can be seen in Figure 11, AD Seniors were significantly slower on the IT task ($F(1,57) = 30.94$, $MS(Error) = 12{,}234{,}349.0$, $p < 0.001$) than Non-AD Seniors. Again reviewing Figure 11, it is clear that, at each level of confidence, the AD Seniors' response times are longer than those of their Non-AD counterparts. Also, as in the Cities in Canada task, the general pattern of decreased RT's is observed as confidence increases.

Also of note is the significant difficulty effect ($F(1,57) = 6.12$, $MS(Error) = 85528.89$, $p < .01$) with participants taking more time to respond on hard items (mean = 2373.64ms) than on easy items (mean = 2235.47ms). This pattern is further observed when the task is broken down by SOA ($F(2,114) = 14.16$, $MS(Error) = 140389.69$, $p < 0.001$) with participants exhibiting slower response times on medium SOAs (mean = 2233.60ms) than on long SOAs (mean = 2217.65ms) and still slower response times on short SOAs (mean = 2462.42ms) than on medium SOAs.

No significant interactions were observed for response time. It is notable, however, in reviewing AD and Non-AD seniors' response times for each SOA (Figure 12) that both groups employ the same underlying decision-making process as the plots for each are almost perfectly parallel.

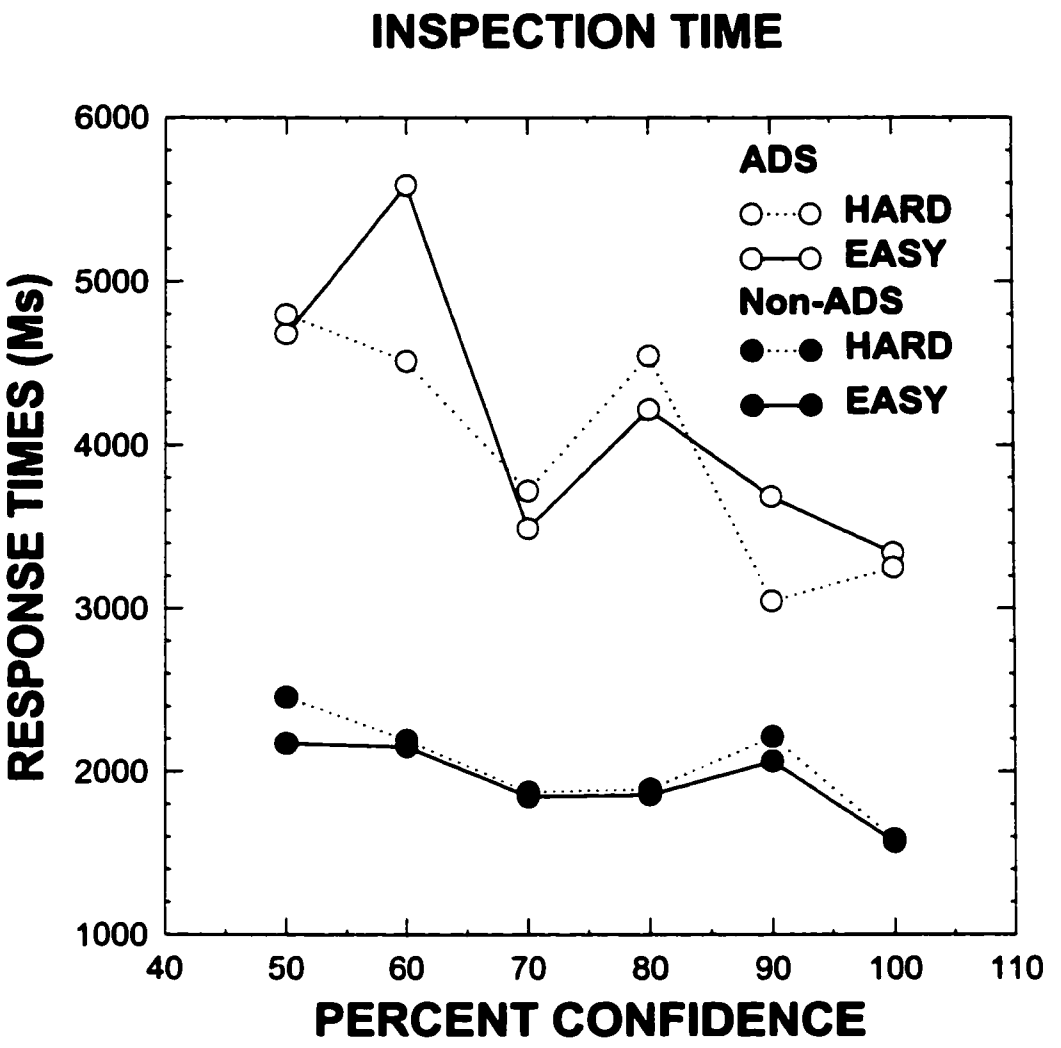


*Figure 11.* Response times by confidence category for Easy items (solid lines) and Hard items (dashed lines) on the Inspection Time task for Seniors who have not been diagnosed with AD (Non-ADS) and Seniors who have been diagnosed with probable or possible AD (ADS).

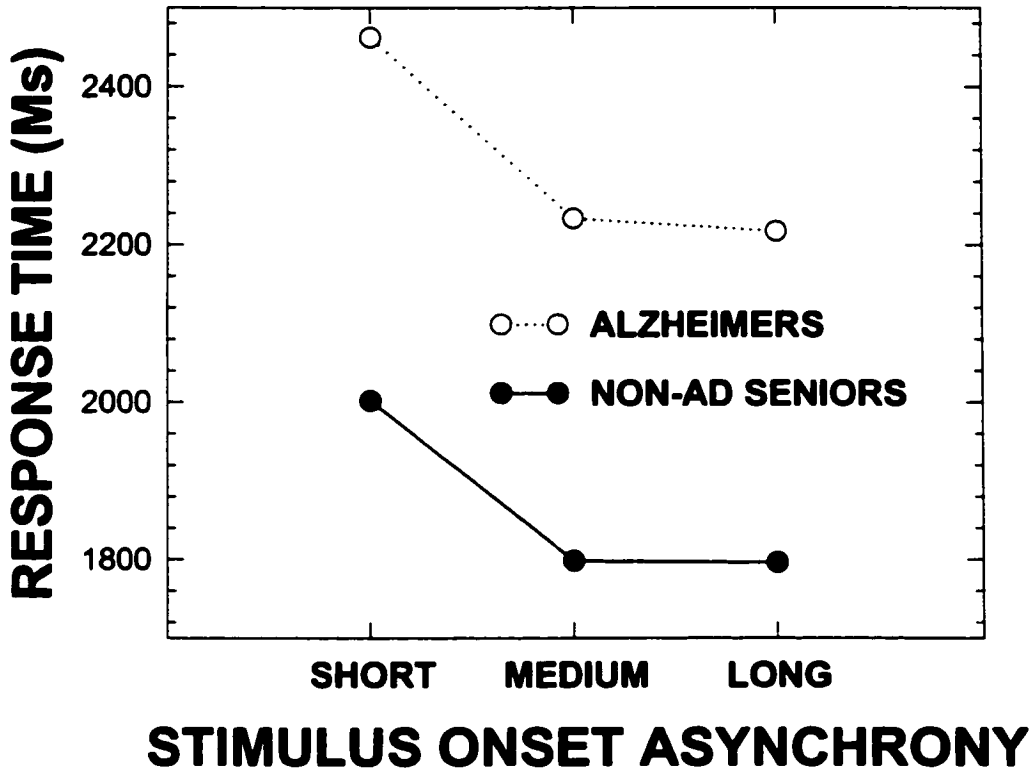


*Figure 12*. Response times for three lengths of stimulus onset asynchrony on the Inspection Time task for AD and Non-AD groups of seniors.

## General Discussion

Of interest in these findings is the existence of over-confidence on hard items for Non-AD Seniors on the Cities in Canada task. This could be argued to be problematic for the PMM model which cannot account for this effect in the general knowledge domain when there is no experimenter bias in item selection. However, it could also be argued that, as pairs were randomly selected within each level of difficulty rather than from across all difficulty level, that experimenter bias did exist in item selection and that these results, therefore, support the PMM model.

Perhaps the clearest effect is the difference in response times between groups. This difference was significant for both the memory and perception tasks. Due to the ability of response time to distinguish between these groups of seniors, a distinction that is made across domains, there may be a role to play for response time in detecting cognitive loss due to Alzheimer's disease.

Also notable is the distinctive pattern seen in the relative frequencies plots for the use of confidence categories on both the Inspection Time task and the Cities in Canada task (see Figures 5, 8 and 9) with AD seniors employing the 50% confidence category more than any other on both easy and hard items despite showing greater accuracy on easy items than hard. On observation, the possible reason for this pattern is also a confound to the current research. As AD seniors endeavoured to complete the tasks for this study, it was necessary to prompt participants after each trial on every task as well as for each confidence decision.

Two things were apparent upon observing the AD seniors during this process. First, it seemed easier for them to understand the concept of the Cities in Canada task and

responded appropriately once cued. It appeared, on the other hand, that the concept of the perceptual task was more difficult for these participants to grasp. A possible reason may be that the type of question asked in the memory task is one that may be encountered normally in one's day-to-day activities. For instance, one can imagine being asked if they happen to know whether Hamilton is larger or smaller than Montreal. Even if the answer is not known, the question itself is likely not to be viewed as unusual. Contrarily, how often, if ever, is one asked to determine whether one briefly presented line is shorter or longer than another? This is not a concept that is easy to grasp and observation of AD seniors suggests that this task was not well understood. Why, then, are the response times for the Inspection Time task so much quicker for the AD participants than their response times on the Cities in Canada task? It could be that, in not understanding the premise of the task, the decision could be made more quickly as they were making a guess right away rather than requiring time to ponder the choice and then responding with a guess.

A number of limitations exist in the current research. First, the difficulty in obtaining sufficient numbers for the clinical sample (AD seniors) resulted in hugely unequal sample sizes. This resulted in reduced power and puts into questions the accuracy of the patterns of results observed. In addition to the difficulties associated with recruitment, computer failure also resulted in the loss of data for a number of AD seniors, further reducing the amount of usable data with which to make comparisons. Ideally, future research should endeavor to attain equal sample sizes of sufficient numbers to obtain adequate power and hence, more conclusive findings.

## Summary

As was expected, difficulty effects were generally observed on both the Cities in Canada memory task and the Inspection Time perception task. Participants were more accurate, had higher mean confidence, were less overconfident or under-confident and better calibrated on easy items compared to hard items. Participants were not significantly better resolved, however, on easy items compared to hard.

Contrary to expectations, there were no group differences on any of the calibration indices for either the Cities in Canada memory task even though, as expected, accuracy was significantly higher in Non-AD seniors compared to AD participants. With respect to resolution, there was a trend towards better resolution for Non-ADS in the memory task but, perhaps surprisingly, towards the AD seniors in the perceptual task.

## Conclusions

Although results were mixed in their support of the hypotheses for this research, interesting findings were obtained nonetheless. Difficulty effects were observed, as expected, and although no group differences were observed as anticipated on the calibration indices for the memory task, neither were they observed for the perception task.

Although results were not strongly supportive of the final hypothesis of this study, other findings are of considerable interest. Group differences in the use of confidence categories indicated an interesting and persistent pattern of relative frequency across domains for the AD seniors as compared to the Non-AD participants. This pattern may have been as a result of regular cuing combined with short-term memory deficiencies. In

future, research designs should account for these factors when looking at the use of confidence categories in AD participants.

Response time data, however, also resulted in consistent group differences without the apparent confounds related to relative frequencies. Response times were significantly slower for AD participants than Non-AD participants on both the memory and perceptual tasks. This indicates the possibility of using simple reaction time data as a means of distinguishing cognitive loss due to Alzheimer's disease from loss due to normal aging, and a direction for future research.

## REFERENCES

Alzheimer Society of Canada (1991). Alzheimer Disease: A Handbook for Care. Toronto, Ontario: The Alzheimer Society of Canada.

Aubin, P. H., & Petrusic, W. M. (1997). Overconfidence in sensory detection. Poster presentation given at the 38th Annual Meeting of the Psychonomics Society, Philadelphia, Pennsylvania, November 20-30.

Backman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease. *Neuropsychologia*, *31*, 335-345.

Baranski, J. V. (1991). Theories of confidence calibration and experiments on the time to determine confidence. Unpublished Doctoral Thesis, Carleton University, Ottawa, Ontario.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgements. *Perception & Psychophysics*, *55*, 412-428.

Baranski, J. V., & Petrusic, W. M. (1995). On the calibration of knowledge and perception. *Canadian Journal of Experimental Psychology*, *49*, 397-407.

Baranski, J.V., & Petrusic, W.M. (1999). Realism of confidence in sensory discrimination. *Perception & Psychophysics, 61,* 1369-1383.

Bjorkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: the underconfidence phenomenon. *Perception & Pscyhophysics*, *54*, 75-81.

Brand, C., & Deary, I.J. (1982). Intelligence and inspection time. In H. J. Eysenck (Ed.), *Models for Intelligence*, pp. 133-148. NY: Springer.

Canadian Study of Health and Aging Working Group (1994). Canadian Study of Health and Aging: study methods and prevalence of dementia. *Canadian Medical Association Journal, 150*, 899-913.

Chui, H. C. (1989). Dementia: A review emphasizing clinicopathologic correlation and brain-behavior relationships. *Archives of Neurology, 46*, 806-814.

Cotrell, V. C. (1997). Awareness deficits in Alzheimer's disease: issues in assessment and intervention. *Journal of Applied Gerontology, 16*, 71-90.

Dawes, R. M. (1980). Confidence in intellectual vs. confidence in perceptual judgments. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice: Papers in honor or Clyde Coombs* (pp. 327-345). Bern: Hans Huber.

Evans, I. A., Funkenstein, H., Albert, M. S., Scherr, P. A., Cook, N. R., Chown, M. J., Hebert, L. E., Hennekens, C. H., & Taylor, J. O. (1989). Prevalence of Alzheimer's disease in a community population of older persons. Higher than previously reported. *Journal of the American Medical Association, 262*, 2551-2556.

Feher, E. P., Mahurin, R. K., Inbody, S. B., & Crool, T. H. (1991). Anosognosia in Alzheimer's disease. *Neuropsychiatry, Neurophysiology, and Behavioral Neurology, 4*, 136-146.

Forbes, W. F., Gentleman, J. F., & Maxwell, C. J. (1995). Concerning the role of aluminum in causing dementia. *Experimental Gerontology, 30*, 23-32.

Giambra, L. M., Arenberg, D., Zonderman, A. B., Kawas, C. & Costa, P. T. Jr. (1995). Adult life span changes in immediate visual memory and verbal intelligence. *Psychology and Aging, 10*, 123-139.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506-528.

Grady, C. L., & Rapoport, S. I. (1992). Cerebral metabolism in aging and dementia. In: J. E. Birren, R. B. Sloane, and G. D. Cohen (Eds.), *Handbook of Mental Health and Aging*, Second Ed., pp. 201-228.

Green, R. C., Clarke, V. C., Thompson, N. J., Woodard, J. L. and Letz, R. (1997). Early detection of Alzheimer Disease: Methods, markers and misgivings. *Alzheimer Disease and Associated Disorders*, *11 (suppl)*, S1-S5.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411-435.

Hasher, L. & Zacks, R. T. (1998). Working memory, comprehension and aging: a review and a new view. In G. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, pp. 193-225. San Diego: Academic Press.

Henmon, V.A.C. (1911). The relation of the time of a judgement to its accuracy. *Psychological Review*, 18, 186-201.

Hodges, J. R, & Patterson, K. (1995). Is semantic memory consistently impaired early in the course of Alzheimer's disease? Neuroanatomical and diagnostic implications. *Neuropsychologia*, *33*, 441-459.

Ishihara, S. (1968). *Tests for Colour-Blindness*: 38 plates Edition. Kanehara Shuppan Co., Ltd., Tokyo, Japan.

Johnson, D. H. (1939). Confidence and speed in the two-category judgement. *Archives of Psychology*, No. 241, 1-52.

Juslin, P., & Olsson, H. (1997). Thurstonian- and Brunswikian origins of uncertainty in judgement: A sampling model of confidence in sensory discrimination. *Psychological Review*, *104*, 344-366.

Juslin, P., Olsson, H., & Bjorkman, M. (1997). Brunswikian and Thurstonian origins of bias in probability assessment: On the interpretation of Stochastic components of judgement. *Journal of Behavioral Decision Making, 10,* 189-209.

Keren, G. (1988), On the ability of monitoring non-veridical perceptions and uncertain knowledge: some calibration studies. *Acta Psychologica*, *67*, 95-119.

La Rue, A., Yang, J., & Osato, S. (1992). Neuropsychological Assessment. In: J. E. Birren, R. B. Sloane, & G. D. Cohen (Eds.), *Handbook of Mental Health and Aging*, Second Ed., pp. 643-670. Toronto: Academic Press.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgements. *Organizational Behavior and Human Performance*, *20*, 159-183.

Lipinska, B., & Backman, L. (1996). Feeling-of-knowing in fact retrieval: further evidence for preservation in early Alzheimer's disease. *Journal of the International Neuropsychological Society*, *2*, 350-358.

Lopez, O. L., Becker, J. T., Somsak, D., Dew, M. A., & DeKosky, S. T. (1994). Awareness of cognitive deficits and anosognosia in probable Alzheimer's disease. *European Neurology*, *34*, 277-82.

MacLeod, C. M. (1992). The Stoop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General*, *121*, 12-14.

Masur, D. M., Fuld, P. A., Blau, A., Levin, H. S., & Aronson, M. K. (1989). Distinguishing normal and demented elderly with the selective reminding test. *Journal of Clinical and Experimental Neuropsychology*, *11*, 615-630.

Mayeux, R. & Schofield, P. W. (1994). Alzheimer's disease. In W. R. Hazzard, E. L. Bierman, J. P. Blass, W. H. Ettinger, Jr., & J. B. Halter (Eds.), *Principles of Geriatric Medicine and Gerontology*, Third Ed., pp. 1035-1050. New York: McGraw-Hill.

McDowd, J. M. & Filion, D. L. (1995). Aging and negative priming in a location suppression task: the long and the short of it. *Psychology and Aging*, *10*, 34-47.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939-944.

Michon, A., Deweer, B., Pillon, B., & Agid, Y. (1994). Relation of anosognosia to frontal lobe dysfunction in Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *57*, 805-809.

Migliorelli, R., Teson, A., Sabe, L., Petracca, G., Petracchi, M., Leiguarda, R., & Starkstein, S. E. (1995). Anosognosia in Alzheimer's disease: a study of associated factors. *Journal of Neuropsychiatry and Clinical Neurosciences*, *7*, 338-344.

Mortimer, J. A. (1983). Alzheimer's disease and senile dementia: Prevalence and incidence. In B. Reisberg (Ed.), *Alzheimer's Disease: The standard reference*. New York: Free Press.

Moss, M. B., & Albert, M. S. (1992). Neuropsychology of Alzheimer's disease. In: r. F. White (Ed.) *Clinical Syndromes in Adult Neuropsychology: The Practitioner's Handbook.* Elsevier Science Publishers, pp. 305-343.

Mullen, R., Howard, R., David, A., & Levy, R. (1996). Insight in Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *11*, 645-651.

Murphy, A. H. (1973). A new vector partition of the probability score. Journal of *Applied Meteriology*, *12*, 596-600.

Nickerson, R. S., & McGoldrick, C. C. (1995). Confidence ratings and the level of performance on a judgmental task. *Perceptual & Motor Skills*, *20*, 311-316.

Norton, L. E., Bondi, M. W., Salmon, D. P., & Goodglass, H. (1997). Deterioration of generic knowledge in patients with Alzheimer's disease: evidence from the Number Information Test. *Journal of Clinical and Experimental Neuropsychology*, *19*, 857-866.

Oakley, K.J. (1998). Age and individual differences in the realism of confidence. Unpublished Doctoral Thesis, Carleton University, Ottawa, Ontario.

Oakley-McKeen, K. J., & Petrusic, W. M. (1995). Calibration and resolution in visual recognition memory. Poster presented at The Canadian Society for Brain, Behaviour and Cognitive Science, Halifax, June 14-19.

Oakley-McKeen, K. & Petrusic, W. (1996). Age effects in the calibration and resolution of confidence in visual detection, size discimination, recognition memory, mental rotation and general knowledge tasks. Poster presentation given at the conference for The Canadian Society for Brain, Behaviour and Cognitive Science (CSBBCS), Montreal, Quebec, August 16.

Olsson, H., & Winman, H. (1996). Underconfidence in sensory discrimination: The interaction between experimental setting and response strategies. *Perception & Pscychophysics*, *58*, 374-382.

Pappas, B. A., Sunderland, T., Weingartner, H. M., Vitiello, B., Martinson, H., & Putnam, K. (1992). Alzheimer's disease and feeling-of-knowing for knowledge and episodic memory. *Journal of Gerontology*, *47*, p159-p164.

Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 962-986.

Petrusic, W. M., & Baranski, J. V. (1989a). Context, context shifts, and semantic congruity effects in comparative judgements. In D. Vickers & P. L. Smith, *Human Information Processing: Measures, Mechanisms and Models*. North-Holland: Elsever Science.

Petrusic, W. M., & Baranski, J. V. (1989b). Semantic congruity effects in perceptual comparisons. *Perception & Psychophysics*, *45*, 439-452.

Petrusic, W. M., & Baranski, J. V. (1997). Context, feedback, and the calibration and resolution of confidence in perceptual judgements. *American Journal of Psychology*, *110*, 543-572.

Pfeffer, R. I., Afifi, A. A., & Chance, J. M. (1987). Prevalence of Alzheimer's disease in a retirement community. *American Journal of Epidemiology*, *125*, 420-436.

Prinz, P. N., & Vitiello, M. V. (1989). Dominant occipital (alpha) rhythm in early stage Alzheimer's disease and depression. *Electroencephalographic Clinical Neurophysiology*, *73*, 427-432.

Raskind, M. A. & Peskind, E. R. (1992). Alzheimer's disease and other dementing disorders. In: J. E. Birren, R. B. Sloane, & G. D. Cohen (Eds.), *Handbook of Mental Health and Aging*, Second Ed., pp. 477-513. Toronto: Academic Press.

Reed, B. R., Jagust, W. J., & Coulter, L. (1993). Anosognosia in Alzheimer's disease: Relationships to depression, cognitive function, and cerebral perfusion. *Journal of Clincial and Experimental Neuropsychology*, *15*, 231-244.

Rocca, W.A., Amaducci, L. A., & Schoenberg, B.S. (1986). Epidemiology of clinically-diagnosed Alzheimer's disease. *Annals of Neurology*, *19*, 415-424.

Selkoe, D. J. (1989). Molecular pathology of amyloidogenic proteins and the role of vascular amyloidosis in Alzheimer's disease. *Neurobiology of Aging*, *10*, 387.

Sharp, G. L., Cutler, B. L. & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgments. *Organizational Behavior and Human Decision Processes*, *42*, 271-283.

Small, G. (1997). Early diagnosis of Alzheimer's Disease. In: L. S. Schneider (Ed.), *Developments in Geriatric Psychiatry,* pp. 39-51. San Fransisco: Jossey-Bass.

Starkstein, S. E., Chemerinski, E., Sabe, L., Kuzis, G., Petracca, G., Teson, A., & Leiguarda, R. (1997). Prospective longitudinal study of depression and anosognosia in Alzheimer's disease. *British Journal of Psychiatry*, *171*, 47-52.

Storandt, M., Botwinick, J., Danziger, W. L., Berg, L., & Hughes, C. P. (1984). Psychometric differentiation of mild senile dementia of the Alzheimer type. *Archives of Neurology*, *41*, 497-499.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-661.

Vasterling, J. J., Seltzer, B., Foss, J. W., & Vanderbrook, V. (1995). Unawareness of deficit in Alzheimer's disease: Domain-specific differences and disease correlates. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *8*, 26-32.

Vickers, D., Nettleback, T., & Willson, R.J. (1972). Perceptual indices of performance: The measurement of "inspection time" and "noise" in the visual system. *Perception*, *1*, 263-295.

Vickers, D. & Pietsch, A. (2001). Decision making and memory: a critique of Juslin and Olsson's (1997) sampling model of sensory discrimination. *Psychological Review*, *108*, 789-804.

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, *110*, 610-617.

Yates, J. F. (1990). *Judgement and Decision Making*, pp. 3-111. New Jersey: Prentice Hall.

**Footnotes**

1. I am grateful to Dr. Petrusic for providing the notation and assistance with this section.

2. Vickers and Pietsch (2001) have provided a forceful and cogent critique of the Sensory Sampling Model. They show that the prediction of uniform underconfidence is an artifact of underlying conceptual and empirical flaws in the model.

3. Unfortunately, neither the demographic nor the neurological data are available. However, at the time of data collection, the information in the medical files corroborated the diagnosis of Alzheimer's disease.

# APPENDIX A


**Form 14**
***Mental Health Act***

**Consent to the Disclosure, Transmittal or Examination of a Clinical Record under Section 29 of the Act**

I, _____
(print full name of person)

of _____
(address)

hereby consent to the disclosure or transmittal to or the examination by KATE OAKLEY, Ph.D.
(print name)

of the clinical record compiled in psychology - Elisabeth Bruyère Pavilion
(name of psychiatric facility)

in respect of /myself ( )
(name of patient) (date of birth, where available)

See Notes 4 and 5.

_____ (witness)

_____ (signature)

_____
(if other than the patient, state relationship to the patient)

Dated the _____ day of _____, 19 ____.

NOTES:

1. Consent to the disclosure, transmittal or examination of a clinical record may be given by the patient where mentally competent or, where the patient is not mentally competent, by the person authorized under section 1*a* of the Act to consent on behalf of the patient. See subsection 29(3) of the Act.

2. Clause 29(1)(b) of the Act provides,
   "(b) 'patient' includes former patient, out-patient, former out-patient and anyone who is or has been detained in a psychiatric facility."

3. Clause 1(g) of the Act provides,
   "(g) 'mentally competent' means having the ability to understand the subject-matter in respect of which consent is requested and able to appreciate the consequences of giving or withholding consent."

4. Subsection 1*a*(1) of the Act provides,
   "1*a*.--(1) A person may give or refuse consent on behalf of a patient who is not mentally competent if the person has attained the age of sixteen years, is apparently mentally competent, is available and willing to give or refuse consent and is described in one of the following paragraphs:
   1. The committee of the person appointed for the patient under the *Mental Incompetency Act*.
   2. The patient's representative appointed under section 1*b* or 1*c*.
   3. The person to whom the patient is married or the person of the opposite sex with whom the patient is living outside marriage in a conjugal relationship or was living outside marriage in a conjugal relationship immediately before being admitted to the psychiatric facility, if in the case of unmarried persons they,
      i. have cohabited for at least one year,
      ii. are together the parents of a child, or
      iii. have together entered into a cohabitation agreement under section 53 of the *Family Law Act, 1986*.
   4. A child of the patient.
   5. A parent of the patient or a person who has lawful custody of the patient.
   6. A brother or sister of the patient.
   7. Any other next of kin of the patient.
   8. The Official Guardian."

## APPENDIX B1

### Informed Consent

The purpose of an informed consent is to ensure that you understand the purpose of the study and the nature of your involvement, in order to determine whether you wish to participate in the study.

**Research Study: Development of a pilot measure for early detection of cognitive loss in Alzheimer's Disease**

Purpose. The purpose of this study is to develop a potentially useful test for early assessment of changes to memory and perception by examining how confident people are when they make choices on certain visual tasks. Perception refers to how well you are able to make sense of what you can see.

Task Requirements: First you will be asked to complete a series of brief visual tests (i.e., test of vision, and a test to determine whether a small cross appears on the left or right of a small line). Then you will be asked to complete a questionnaire about your self-confidence and provide personal information (e.g., age, gender, education). Depending upon your vision, you may be invited to participate in several tests of memory and perception. Each task will ask you to make choices about items projected onto a movie or computer screen, and then to indicate how confident you feel about that choice. For example, tasks consist of reporting whether you see a gap in a centre of a line; choosing the larger or smaller of two line-lengths; choosing the larger or smaller of two squares; choosing the larger or smaller population of two cities; determining whether two rotated "L" shapes are the same or different; or identifying which item from a set of four pictures you were shown. You will be given plenty of opportunity to practice each task. While some tasks are fairly easy, other tasks are extremely difficult for everyone and finding the tasks difficult should not indicate a deficit. Upon request, results of the study can be provided as soon as the study is completed and data analyzed for all participants.

Duration: Each task takes approximately 20-25 minutes to complete, for a total of approximately 3 to 4 hours. The tasks are completed over several short sessions, and rest breaks are provided. You will be reimbursed for any out-of-pocket expenses associated with your participation in this study, such as a light meal or parking.

Potential Risk/Discomfort. There are no potential physical or psychological risks in this study other than possibly being tired.

Anonymity/Confidentiality. All data collected in this experiment are confidential, and are numerically coded such that your name is not associated with the data. Your name and other identifying information will not appear on data sheets, and coded data are made available only to the researchers associated with this project. Also, results presented at conferences and in professional journals or to the funding agency will only be presented in aggregate form or using the number code such that no single person can be identified.

Right to Withdraw. You have the right to refuse to participate in any specific task or the entire project. You also have the right to withdraw your consent and terminate your participation in this project at any time for any reason. Your withdrawal will not affect your continuing or future medical care in any way.

Research Personnel. If you have any questions or comments, please contact either:

Dr. K. Oakley (Principal Investigator), at 562-0050, ext. 1230
Dr. J. Kozak, Director, Research Department, at 562-6365
Dr. P. Soucie, President, SCOHS Research Ethics Board, at 562-0050

---

I have read the above description of "Development of a pilot measure for early detection of cognitive loss in Alzheimer's Disease" and understand the conditions of my participation. My signature indicates agreement to participate in this study, to allow the researcher access to any relevant test results for statistical purposes only, and to confirm that I have received a copy of this consent form.

| | | |
|---|---|---|
| Participant's Name | (Signature) | (Date) |
| Substitute Decision-Maker (if applicable) | (Signature) | (Date) |
| Name of Witness | (Signature) | (Date) |

# APPENDIX C

## CHOICE RESPONSE TIME (CRT) - Instructions

In this task, you will see a small "cross" in the centre of the screen. Next, a small line will appear either to the left of the "cross" or to the right of the "cross". Your task is to decide whether the small line is on the left or the right of the cross. If you think the line is on the left of the cross, press the "left" button marked "L"; if you think the line is on the right of the cross, then you would press the "right" button marked "R".

First, we'll do some practice trials. Then you will be presented with four sets with a short break in-between. Press any key on the keyboard when you are ready to resume. The task will take approximately 5 minutes. Are there any questions before we begin?

# APPENDIX D

## STROOP - Instructions

In this task, you will presented with either the word "red", the word "green", or three stars ***. They will be presented in one of two colours: red or green. So the *word* "green" could be coloured either red or green. Also, the word "red" or the three stars could be either red or green. Your task is to decide on the *colour* of the word or stars. If you think the word or stars are red, press the "left" button marked "R" for "red"; if you think that the word or stars are green, then you would press the "right" button marked "G" for "green".

First, we'll do some practice trials. Then you will be presented with four sets, with a short break in-between. Press any key on the keyboard when you are ready to resume. The task will take approximately 5 minutes. Are there any questions before you begin?

## APPENDIX E

SERVICE DE SANTÉ *des Soeurs de la Charité d'Ottawa inc.*
*Sisters of Charity of Ottawa* HEALTH SERVICE INC.

### DEVELOPMENT OF A PILOT MEASURE FOR EARLY DETECTION OF COGNITIVE LOSS IN ALZHEIMER'S DISEASE

I am writing to invite you to participate in an important research project exploring the early assessment of changes to memory and perception. In this regard, I would like to introduce you to Dr. Kate Oakley who is conducting this study at our hospital. She is examining how confident people are when they make decisions about visual information. Your participation may help provide important information that will lead to early detection of Alzheimer's Disease.

Dr. Oakley needs individuals who have experienced problems with their memory, as well as individuals who have *not* had any memory problems, to participate in this study. She will contact you shortly by telephone to provide more information about this project. This contact does not mean that you are under any obligation to participate, although I encourage you to do so. Should you wish to contact Dr. Oakley, she can be reached at 562-0050, ext. 1230.

Sincerely,

D.A. Guzman, M.D., FRCPC
Director, Memory Disorder Clinic

## APPENDIX F

### GENERAL KNOWLEDGE OF CITIES - Instructions

In this task, you will be presented with pairs of names of two cities. Your task is to decide which of the two cities has either the "larger" population or the "smaller" population, depending upon the instruction. The instruction "larger" or "smaller" will be presented at the top of the screen in the centre. If you think that the city on the left corresponds to the instruction, you are to press the button on the response panel on the left, marked "L". If you think that the city on the right corresponds to the instruction, then press the button marked "R" on the right. Please respond as quickly and as accurately as you can.

Next, you will be asked to record your record your level of confidence about your decision, using the response panel, where 50 means a "guess", and the 60-90 buttons indicate increasing confidence in your decision, with the 100 button reserved for decisions where you are "completely certain" of your response.

If you pressed the wrong button when making your choice, and you know you made a mistake, then you would use the button marked 0 to indicate a "mistake".

There is only one part for this task, but you are provided with a short 'rest' period approximately half way through the task. This task takes approximately 20-25 minutes to complete.

You will first receive some practice trials. Are there any questions before you begin the practice session?

## APPENDIX G

### INSPECTION TIME - Instructions

In this task, you will be presented with a pair of lines connected by a bar (show example).

The pair of lines will be presented very briefly, and covered over by another line-pair that is thicker and longer, which will mask the first pair of lines (show example).

Your task is to decide which of the first two lines is either the "shorter" or the "longer" of the two. The instruction "shorter" or "longer" will be randomly presented at the top of the computer screen. Indicate your choice on the response panel, where the "left" button indicates that you think the line on the left corresponds to the instruction, and the "right" button indicates that you think the line on the right corresponds to the instruction. Please try to be as *accurate* as possible.

Next, you will be asked to record your level of confidence about your decision, using the response panel, where 50 means a "guess", the 60-90 buttons indicate increasing certainty of your decision, and 100 means completely "certain" of your decision.

If you pressed the wrong button when making your choice, and you know you made a mistake, then you would use the button marke 0 to indicate a "mistake".

First, you will receive a practice session. I will stay with you during the practice session in case you have any questions. OK, are there any questions before we begin?

EXPERIMENTER: At the conclusion fo the practice block, please ask the subject to identify any strategies that they may have used in making a decision about the line lengths (i.e., movement, one line jumps), and record on the back of the index card

At the end of each set, you will be asked to indicate whether you saw any movement in one of the lines that you used to help you make a decision. Some people report seeing movement, and others do not. Each individual is different... it depends upon your visual ability. DID YOU SEE ANY MOVEMENT? If you did, please enter the proportion or percentage of trials on which you observed movement (from 1-100%) using the computer keyboard. (Experimenter: make sure 'numlock' is depressed). Next you will be asked to indicate the amount of that movement, on a scale from 1-9. If you did not see any movement, please enter "0".

Now for the actual task. You will be presented with two sets. Following this response, you will be provided with a short break before going on to the next set. When you are ready to resume the task, simply press any key on the keyboard to continue. This task takes approximately 20-25 minutes.