

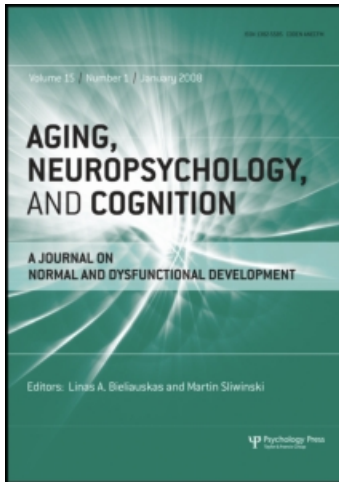
This article was downloaded by: [University of Plymouth]

On: 3 April 2009

Access details: Access Details: [subscription number 909915483]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Aging, Neuropsychology, and Cognition

Publication details, including instructions and subscription information:

<http://www.informaworld.com/smpp/title-content=t713657683>

Global Predictions of Memory in Alzheimer's Disease: Evidence for Preserved Metamemory Monitoring

Chris J. A. Moulin; Timothy J. Perfect; Roy W. Jones

Online Publication Date: 02 December 2000

To cite this Article Moulin, Chris J. A., Perfect, Timothy J. and Jones, Roy W.(2000)'Global Predictions of Memory in Alzheimer's Disease: Evidence for Preserved Metamemory Monitoring',Aging, Neuropsychology, and Cognition,7:4,230 — 244

To link to this Article: DOI: 10.1076/anec.7.4.230.796

URL: <http://dx.doi.org/10.1076/anec.7.4.230.796>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Global Predictions of Memory in Alzheimer's Disease: Evidence for Preserved Metamemory Monitoring*

Chris J.A. Moulin^{1,2}, Timothy J. Perfect³, and Roy W. Jones²

¹University of Reading; ²Research Institute for the Care of the Elderly, St. Martin's Hospital, Bath, UK; and ³University of Plymouth

ABSTRACT

Previous researchers have argued that there is a metamemory monitoring deficit in Alzheimer's disease (AD) because patients tend to overestimate their recall performance on a word list. We propose that these previous results are a misleading by-product of the methodologies used, rather than evidence of an underlying metamemory deficit. In two experiments, AD patients and older adult controls made predictions of performance both before and after encoding a to-be-remembered list. Metamemory function was measured by observing the shift in predictions made with, and without, an opportunity to monitor the list. Experiment 1 found that although there were differences between the groups' accuracy for their prestudy predictions of recall, both groups were equally accurate after encoding. Experiment 2 explored this using four lists that varied in item difficulty and semantic relatedness. This experiment replicated the findings of Experiment 1, and it was also found that the AD group became more accurate at predicting their performance with more exposure to study-test trials. These studies suggest that metamemory monitoring is intact in AD, because AD patients utilize information gained during processing the to-be-remembered items to revise their predictions of subsequent performance.

Metamemory is comprised of monitoring and control processes (Nelson & Narens, 1990). When items are being encoded in an episodic memory task, cognitive resources are both monitored and controlled. For example, an item that is considered difficult to remember (by monitoring) may be studied for longer (by controlling). The focus of this paper is metamemory monitoring. In particular we propose that if metamemory is intact in Alzheimer's disease (AD), we will be able to observe participants revising their metamemory judgements as a result of feedback from memory monitoring.

A deficit in metamemory has been put forward as one factor that might contribute to episodic memory dysfunction (e.g., Light, 1991; Shimamura & Squire, 1986). Researchers aiming to explain the episodic memory loss experienced in AD have adopted a metacognitive framework, in particular examining memory monitoring (e.g., Correa, Graves, & Costa, 1996; McGlynn & Kaszniak, 1991). One way in which monitoring proficiency has been measured is to ask participants to make an estimate of the number of items from a list they will recall. These judgments of list performance have

* The authors would like to thank the staff and patients of the RICE (Research Institute for the Care of the Elderly, St. Martin's Hospital, Bath) and BRACE (Bristol Research into Alzheimer's Disease and Care of the Elderly, Blackberry Hill Hospital, Bristol) memory clinics. This work was supported by an Economic and Social Research Council studentship award to the first author, and was carried out at the Department of Experimental Psychology, University of Bristol, UK. Parts of this research were presented at the Experimental Psychology Society (London, UK, 1998) and Cognitive Aging (Atlanta, GA, 1998) Conferences.

Address correspondence to: Chris Moulin, Department of Psychology, The Harry Pitt Building, University of Reading, Earley Gate, Reading, RG6 6AL, UK. Tel.: 44 (0) 118 931 8523. Fax: 44 (0) 118 931 6715. E-mail: c.j.a.moulin@reading.ac.uk

Accepted for publication: July 17, 2000.

been labeled *global predictions* (e.g., Connor, Dunlosky, & Hertzog, 1997). Using evidence from global predictions, it has been argued that AD patients display deficient memory monitoring abilities and this compounds their episodic memory problems (McGlynn & Kaszniak, 1991). Within the AD literature, these global judgments have either been made before study (known as predictions, e.g., Schacter, McLachlan, Moscovitch, & Tulving, 1986) or after test (known as postdictions, e.g., Correa et al., 1996). They have been made for a variety of tasks: categorized word lists (Schacter et al., 1986); everyday tasks and a battery of cognitive tests (McGlynn & Kaszniak, 1991); and cued recall and verbal learning tests (Correa et al., 1996). Without exception, these studies report that AD patients overestimate their performance. From this it has been inferred that memory monitoring is impaired in AD, even relative to other memory impaired groups (e.g., AD vs. age-associated memory impairment; Feher, Larrabee, Sudilovsky, & Crook, 1994).

However, one aspect of memory monitoring is overlooked in these studies of AD: the pattern of participants' judgments before and after study. It can be argued that although the AD group overestimates their performance, they may still be monitoring memory accurately. That is, AD patients may make judgments that are reflective of their cognitive processes at encoding, but which are poorly calibrated with regards to final recall. This argument resonates with evidence from the normal aging literature. Connor et al. (1997) present a review of previous research into global predictions in aging. Some studies in the nondemented old suggest that older adults accurately monitor their performance (e.g., McDonald-Miszczak, Hunter, & Hultsch, 1994; Rebok & Balcerak, 1989), whereas other studies suggest that older adults are not accurate (Bruce, Coyne, & Botwinick, 1982; Perlmutter, 1978). As with the AD literature, older adults tend to overestimate memory performance in the predictions that they make (e.g., Bruce et al., 1982; Devolder, Brigham, & Pressley, 1990), although some studies indicate that they underestimate it (e.g., McDonald-Miszczak et al., 1994).

One explanation of this pattern has been suggested by Connor et al. (1997). They propose that these inconsistencies could be due to the use of the midpoint of the scale for memory performance. This suggests that individuals "anchor" their predictions near the midpoint of the possible range of performance – treating it as an "intuitively plausible performance level" (p. 51), considering that they know little about the memory task they are about to undertake. Connor et al. argue that overestimating performance does not necessarily reflect a metamemory deficit in the old, because depending on the length of the to-be-remembered list, normal older adults show the same pattern of performance as younger participants. Therefore, Connor et al. emphasize the change in predictions from participants at different stages during the memory task. From the relative pattern across judgments it is possible to observe memory monitoring. According to Nelson and Narens' (1990) framework, participants who alter their predictions according to exposure to the to-be-remembered list are basing their judgments on memory monitoring. Connor et al. demonstrated that both the young and old became more accurate at predicting performance between a prestudy prediction and a poststudy prediction. In conclusion, although there were differences in absolute levels of judgment accuracy due to the midpoint anchoring effect, older adults still show the same sensitivity in that they revise their predictions following study, and their postdictions are more accurate than their predictions.

In the present study we adopt the same emphasis on the change in predictions to ascertain whether AD patients' estimates of performance are sensitive to whether they have studied the list or not. There is some evidence that despite overestimating performance, AD participants do monitor performance. For example, Lopez, Becker, Somsak, Dew, and DeKosky (1994) investigated anosognosia (unawareness of deficit) in 181 AD patients. Firstly, before testing, Lopez et al. asked: "Do you have a memory problem that makes your everyday life more difficult or complicated?" (p. 278). From this, 44% of patients denied that they had any memory problems. After this question, participants

were administered a standardized neuropsychological assessment, after each stage of which they were asked by the examiner, "How do you think you did those tasks?" (p.278). Participants who spontaneously complained of abnormal performance or who agreed when questioned further about their difficulty were considered aware of deficits. At this stage, after testing, the number of participants who were unaware was reduced to 23%. Therefore, in Lopez et al.'s study, participants were more likely to be aware of their deficit having completed cognitive testing. This was not compelling evidence of feedback from monitoring cognitive processes, however, because Lopez et al. suggest that theirs was a rather "crude" measure without any scaling to allow the measurement of degree of awareness. Furthermore, it was not clear how the insistent and explicit questioning about awareness affected the participant's judgment, or what control performance would have been.

More evidence that AD patients are monitoring memory comes from findings that AD patients' judgments are sensitive to the task. McGlynn and Kaszniak (1991) found that the AD group's postdictions were significantly more discrepant than controls for some tasks (word recall, picture recall, delayed picture recognition, digit span, verbal span, and verbal fluency). However, there were no differences between groups on other, arguably easier, tasks (word recognition, immediate picture recognition, and spatial span). McGlynn and Kaszniak did not supply information on the relative levels of performance for these tests but they proposed that participants particularly overestimate their performance on tasks that have dramatically worsened as a result of their dementia.

One problem with these types of study, as McGlynn and Kaszniak point out, is that there are significant differences between groups on some tests because, whereas the AD group overestimate slightly, the control group underestimate their performance by up to 40%. Often this problem is worsened by researchers using directional discrepancies as measures of accuracy. These are calculated as the estimate of performance minus the actual performance (e.g., Correa et al., 1996). This method makes it im-

possible to interpret group means. For example, two participants who respectively over- and underestimate by four items will have a group mean discrepancy of zero. In contrast, a pair who both overestimate by two will have a mean discrepancy of two, although they are clearly more accurate in their predictions.

Correa et al.'s (1996) study is directly relevant to the present paper. They argue that AD patients have a memory monitoring deficit by focusing on results from a directional discrepancy measure. However, when they analyzed their data using an unsigned absolute difference score (a nondirectional discrepancy), there was no significant group difference between the AD group and a control group (Correa et al., 1996). From this, they argued that a signed difference measure was a more sensitive measure for detecting group differences in estimations of performance. The correct conclusion would have been that the AD group more consistently overestimated performance on this given task.

Here we use nondirectional discrepancy measures (e.g., Hertzog, Saylor, Fleece, & Dixon, 1994), calculated as the modulus of the difference between the prediction and actual performance. They are ideal for the present work because we are interested in seeing whether participants are more accurate after presentation of the list, and the main focus is how accuracy changes after participants have encoded the items. This measure is adequate for examining groups' metamemory sensitivity without the confound of differences in memory performance.

In summary, we adopted Connor et al.'s (1997) emphasis on the pattern of accuracy across different phases of list learning. In the experiments presented here, participants are asked to predict their performance on a list of items, before and after study. By comparing these predictions we can infer whether participants accurately monitor their learning of the to-be-remembered items. Although some studies have measured prestudy predictions and others have used postdictions, to our knowledge no studies have examined global judgments at different stages of the same task in AD. As demonstrated with older adults (Connor et al., 1997), it is conceivable that although AD patients are ini-

tially unaware of their deficit, they do nonetheless monitor performance during an episodic memory test. That is, we aim to ascertain whether AD participants are sensitive to processes occurring during encoding and thus whether they show evidence of memory monitoring in their patterns of prediction. Such a preservation in sensitivity to factors operating at encoding has been shown for item judgments in AD (Moulin, Perfect, & Jones, 2000b). By using a discrepancy measure and comparing predictions at two time points we can make inferences about memory monitoring processes that occur during encoding without the possible confound of subsequent memory performance. Thus, the only way to examine memory monitoring during encoding in a memory impaired group such as AD is to consider the revision of predictions between two time points (Experiment 1) or across a succession of lists (Experiment 2).

EXPERIMENT 1

METHOD

Participants

There were 32 participants, the characteristics of whom are shown in Table 1. Sixteen were patients with a diagnosis of probable or possible AD (McKhann et al., 1984). These patients were recruited from two hospital-based memory clinics and had diagnoses made by independent clinicians. Patients were diagnosed as being demented with

the *Diagnostic and Statistical Manual of Mental Disorders*, third edition (DSM III-R; American Psychiatric Association, 1987) criteria and as having AD by the National Institute of Neurological and Communicative Disorder and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 1984). AD was diagnosed by a clinician using neuropsychological examination, Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), family interview, laboratory screening (i.e., hematology; B12 and folate levels; renal, liver and thyroid function; calcium and syphilis serology), and medical examination. If there was a suggestion of a psychiatric disorder, patients were also assessed by a psychiatrist. Patients with a history of stroke or depression were excluded from this study. Patients with a Hachinski score (Hachinski et al., 1975) that indicated they might have a vascular component to their dementia were also excluded.

Sixteen participants made up an older adult control (OAC) group. These people were either caregivers of patients tested at the memory clinics ($n = 6$) or people recruited from a panel of older adults who had expressed an interest in participating in research ($n = 10$). All OAC participants had recently been screened for dementia using the MMSE, all scoring above the standard cutoff of 26/30. The OAC participants who were part of the volunteer panel received a small remuneration for their time. A one-way ANOVA showed that there was no significant difference in the education level (years of formal education) of these two groups, $F(1, 30) = 1.62$, $MSE = 9.03$. However, the AD group was significantly older than the OAC group, $F(1, 30) = 9.62$, $MSE = 25.71$, $p < .01$. Therefore, the principal analyses were run with age as a

Table 1. Participant Characteristics for Experiments 1 and 2: Mean (Standard Deviation) Age, MMSE, and Education Level.

	Age		MMSE		Education level	
	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)
Experiment 1						
AD	75.75	(5.64)	21.81	(3.10)	9.94	(1.77)
OAC	70.19	(4.43)	–		11.00	(2.83)
Experiment 2						
AD	78.13	(5.29)	17.38	(6.17)	10.88	(2.70)
OAC	75.56	(4.55)	–		11.06	(2.08)

Note. Education level = years of formal education; AD = Alzheimer's disease; OAC = older adult control; MMSE = Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975).

covariate in the analyses of variance (ANCOVA). The effect of age failed to account for any of the differences between groups in all the analyses presented here.

Stimuli/Materials

The word lists comprised 5 easy and 5 difficult items selected from Rubin and Friendly's (1986) recall norms (see Appendix). Words were presented on a computer screen, one at a time. They were 2 cm high in black on a white background. Participants checked similar words for readability before the trial commenced. Nobody reported any problems with readability.

Design and Procedure

In this paper we report only the global judgments of performance. As well as these global judgments, item judgments were made. These are reported elsewhere (Moulin et al., 2000b). This experiment comprised two separate tasks that were both carried out by the same participants. There was a recall readiness task (List 1) and a judgment of learning (JOL) task (List 2). In List 1 participants studied each word for an unlimited time. In List 2 presentation time was fixed (2 s per word) and participants had to give each word a rating of how likely they were to subsequently recall it. These tasks only vary in the different dependent variables taken for the items. In both tasks, the global judgment procedure was identical. Every participant completed the recall readiness task (List 1) before the JOL task (List 2). This combined item and global judgment procedure is used elsewhere in the literature (e.g., Connor et al., 1997).

Participants were tested individually in a quiet room and were first familiarized with the computer and screen. Instructions were presented on the computer screen and also read aloud by the experimenter. Participants were told before the presentation phase that there would be an immediate free recall test, with an unlimited time to recall. They were instructed that there would be 10 items on each list to remember. Participants were told that the aim of the experimental procedure was to correctly predict the number of items they would recall rather than to set a target number of items. However, they were also told that it was acceptable to recall more or less words than they had predicted. Before the presentation phase, participants made a prestudy prediction of their performance (as a figure out of 10). The to-be-remembered items were then presented individually in a random order. Items were presented until the participant declared recall readiness (List 1) or masked after 2 s presentation and a JOL made (List 2). After pre-

sentation of the 10 words, there was a short digit span task (4 digits) to remove short-term memory based recency effects. The poststudy prediction was made following the digit span task. Participants were instructed not to count up the words currently recallable, but rather to intuitively and quickly predict their level of recall. Recall immediately followed the poststudy prediction. All participants responded orally when giving their predictions and recalling items.

RESULTS

The left panel of Figure 1 shows the mean prestudy predictions, poststudy predictions, and actual recall for both lists. Firstly we consider recall performance. There was the expected recall difference between these groups, $F(1, 30) = 92.72$, $MSE = 2.38$, $p < .001$, with the AD group performing significantly worse; however, there was no effect of list, $F(1, 30) < 1$, nor a significant interaction, $F(1, 30) = 1.43$, $MSE = 1.32$.

Figure 1 shows that the AD patients tend to overestimate their recall performance, but that they revise their predictions downward after encoding. To analyse accuracy before and after presentation of the to-be-remembered lists, a nondirectional discrepancy was calculated for each participant (Table 2). A $2 \times 2 \times 2$ (Group \times List \times Judgment Type) ANOVA was carried out on the discrepancy data. There was a main effect of group, $F(1, 30) = 6.38$, $MSE = 6.35$, $p < .05$, indicating that the AD group was less accurate overall. The main effect of list was also significant, $F(1, 30) = 4.68$, $MSE = 4.17$, $p < .05$, and the means indicate that participants were more accurate in the second task. The Group \times List interaction was not significant, $F(1, 30) < 1$, indicating that both groups are more accurate on the second list. The effect of judgment type was significant, $F(1, 30) = 14.85$, $MSE = 1.90$, $p = .001$, as was the Group \times Judgment Type interaction, $F(1, 30) = 9.49$, $MSE = 1.90$, $p < .01$. The means suggest that all groups make more accurate poststudy predictions than prestudy predictions. Simple main effects explored the Group \times Judgment Type interaction. There was an effect of judgment type in the AD group, $F(1, 30) = 24.03$, $MSE = 1.90$, $p < .001$, but not in the

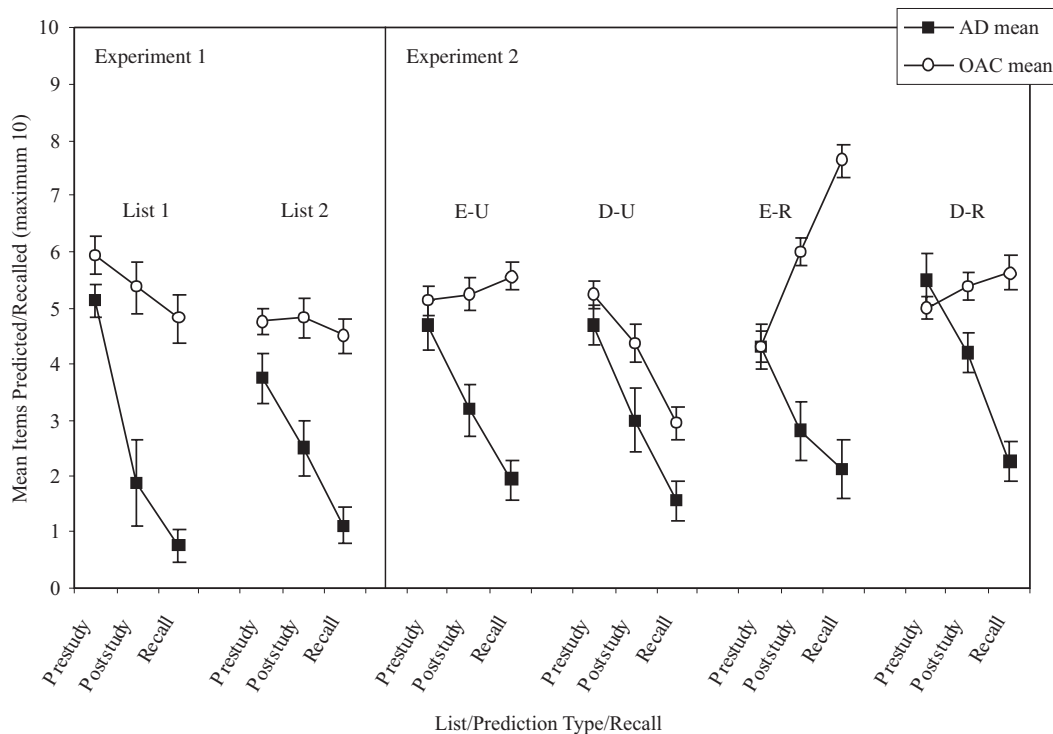


Fig. 1. Mean prestudy and poststudy predictions and actual performance: Experiment 1 and Experiment 2. Error bars = 1 standard error. E-U = easy unrelated list; D-U = difficult unrelated list; E-R = easy related list; D-R = difficult related list.

OAC group, $F < 1$. This indicates that only the AD group revises its predictions by a significant amount once they have been presented the list. Simple main effects also examined group differences for each judgment type. There was only a group difference in accuracy for prestudy predictions, $F(1, 30) = 18.60$, $MSE = 3.02$, $p < .001$. For poststudy predictions, the groups did not differ in their discrepancies, $F < 1$. This indicates that whereas the AD group is initially more discrepant, it revises its estimates to levels which are as accurate as controls. The Trial \times Judgment Type interaction approached significance, $F(1, 30) = 3.77$, $MSE = 1.87$, $p = .062$, and the three-way interaction also approached significance, $F(1, 30) = 3.77$, $MSE = 1.87$, $p = .062$.

To summarize, there is evidence that the AD group is monitoring its memory performance. The AD group revises its estimates of perfor-

mance to more accurate levels having encoded the list, and is more accurate in List 2. Moreover, there is no significant difference in accuracy between groups for the poststudy predictions.

DISCUSSION

As with previous work (e.g., Correa et al., 1996; McGlynn & Kaszniak, 1991), this study found that AD participants significantly overestimate their memory performance. However, this study investigated memory monitoring by comparing predictions before and after study. We found that AD patients make significantly more accurate predictions of performance after study. Therefore, they may be unaware of their recall deficit in absolute terms, but they do display

Table 2. Accuracy of Participant's Predictions of Performance Before and After List Presentation: Mean (Standard Deviation) Nondirectional Discrepancies.

		AD				OAC			
		Prestudy		Poststudy		Prestudy		Poststudy	
		<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)	<i>M</i>	(<i>SD</i>)
Experiment 1	List 1	4.37	(1.63)	1.75	(2.88)	2.12	(2.06)	1.93	(2.32)
	List 2	2.75	(1.69)	2.00	(1.78)	1.25	(0.85)	1.06	(1.06)
Experiment 2	List 1	4.31	(2.36)	2.00	(2.28)	2.12	(1.15)	1.06	(0.92)
	List 2	2.87	(1.96)	1.56	(1.31)	2.25	(1.77)	1.19	(0.98)
	List 3	3.06	(1.87)	1.50	(1.86)	1.81	(2.04)	1.44	(1.09)
	List 4	1.94	(1.26)	1.50	(1.26)	2.12	(0.96)	1.68	(1.30)

Note. AD = Alzheimer's disease ($n = 16$); OAC = older adult controls ($n = 16$); prestudy = prediction made before study; poststudy = prediction made after study; nondirectional discrepancy = modulus of prediction minus actual recall.

evidence of memory monitoring: They are sensitive to factors operating during encoding.

One interesting finding is that AD participants become more accurate from one list to the next, suggesting that they can obtain useful feedback from the test procedure. Here we propose that this increased accuracy across lists, like the benefit of making a poststudy, as opposed to a prestudy prediction, is a result of AD participants being able to monitor their memory during presentation. This conclusion is based on Nelson and Naren's (1990) model. In particular, we suggest that by attempting to encode the items, participants' estimation of their predicted level of performance is altered. Within this metamemory framework, such a change in predictive accuracy can only be a consequence of intact metamemory functioning. Therefore, participants are more accurate in List 2 than in List 1 because of exposure to the study-test procedure.

An important consideration is that this interpretation is confounded by the task that the participants were carrying out while encoding the list. It could be that prompting participants to make JOLs for the items increases their accuracy for the prediction for the whole list (List 2). Experiment 2 will take global judgments only, removing the possible confound of taking simul-

taneous item judgments. In addition, the next experiment keeps tasks constant and counterbalances the order of lists to examine the effect of list position more carefully.

EXPERIMENT 2

Experiment 2 had three aims. Firstly, we removed the possible confound of taking item judgments at the same time as global judgments. Secondly, we wanted to test the hypothesis that through exposure to study and test trials, the AD group become more accurate in their predictions of performance (as suggested in Experiment 1). Thirdly, in Experiment 2 we varied the difficulty and relatedness of the lists to see if modulating list difficulty on a list resulted in differences in global predictions. Because recall was equal on both lists of Experiment 1, it was not known whether participants could vary their predictions according to the different levels of recall for the lists. Presenting participants with objectively different lists enabled the measurement of sensitivity to list differences. If there was a shift in participants' predictions according to objective qualities of the list, we would be able to conclude that participants were sensitive to the quality of the to-be-remembered stimuli. Again,

this would be evidence for intact metamemory functioning that is independent of the effects of poor recall.

Because we argued that the increased accuracy shown in the poststudy predictions in Experiment 1 was due to participants being able to monitor memory, we also expected feedback from monitoring to be sensitive to other factors that affect performance. We were therefore interested in whether the AD group was aware of the characteristics of the list as they were encoding it.

This experiment examined sensitivity to semantic relatedness. A prominent theme in AD research is that there is a deficit in semantic memory (e.g., Nebes, 1989) which may contribute to the episodic deficit (Goldblum et al., 1998). Typically, AD patients do not benefit from semantic relatedness in their recall on episodic tasks. For example, Herlitz and Viitanen (1991) suggest that there is evidence that AD patients “show a deficit in utilizing semantic knowledge as an aid for episodic remembering” (p. 571). Previous research into metacognition in AD has suggested that control and monitoring processes dissociate in AD (Moulin, Perfect, & Jones, 2000a). Therefore, it is conceivable that AD patients are aware of the differences between the lists, but cannot benefit from these differences. Because previous research investigating this issue has focused on output measures (i.e., recall performance; Herlitz & Viitanen, 1991), it is not possible to draw conclusions about how AD participants respond to semantically grouped lists during encoding. It is conceivable that during encoding the AD patients are aware of semantic relatedness and its benefits, but then catastrophic forgetting intervenes. Conversely, the AD group may fail to benefit from the semantic relatedness of lists if they fail to monitor the items, and therefore do not implement the strategies which semantic relatedness facilitates. Taking metamemory measures at encoding allowed us to discriminate between these two possibilities.

Finally, to ascertain whether AD patients become more accurate across the list as a result of exposure to the test procedure, lists were presented in a counterbalanced order to allow a sep-

arate analysis of accuracy across list order. The extent to which AD patients monitor their performance will be reflected in the pattern of their discrepancy across lists. If the AD group benefit from exposure to multiple trials of the same task, we expect them to become less discrepant in their predictions of performance across list order.

METHOD

Participants

There were 32 participants, 16 in each group of AD and OAC. Participants were not the same sample as that of the previous experiment. The AD patients were drawn from the two memory clinics used in Experiment 1, and were recruited according to the same criteria. The OAC participants were recruited from a panel of healthy older adults who had expressed an interest in participating in research, and had been screened on the MMSE as in Experiment 1. There was no significant difference between the groups' mean ages (ANOVA), $F(1, 30) = 2.16$, $MSE = 24.32$, and no significant difference in education level, $F < 1$.

Design and Materials

Again, participants were given lists of 10 words. Relatedness and difficulty of recall were manipulated over four lists in a 2×2 format. The four lists of words were: easy unrelated (E-U), difficult unrelated (D-U), easy related (E-R), and difficult related (D-R). (See Appendix.) There were two levels of relatedness: related in which all the items came from the same semantic category (Battig & Montague, 1969), and unrelated in which none of the items came from the same semantic category. With the related lists, the difficulty of recall for items was manipulated by manipulating the items' typicality. The E-R list used 10 words that were all units of time (e.g., week, day), whereas the D-R list consisted of color names and was made more difficult by using less frequent and more obscure examples of the category (e.g., magenta, olive). For the unrelated lists, words were chosen from Rubin and Friendly's (1986) free recall norms. Difficulty of recall for items was manipulated by using the normative values of the words – one list consisted of 10 words with a high probability of free recall (E-U) and the other list consisted of a set of items with a low probability of free recall (D-U).

Lists were not fully counterbalanced due to the large number of patients that would be required. Instead, all lists were presented in the same order but the starting position was rotated, each participant starting on one of the four lists, but proceeding through the same cycle: E-U, D-U, E-R, D-R. The metamemory judgments were prestudy predictions and poststudy predictions as taken in Experiment 1.

Procedure

Participants were tested individually and were instructed that they would be given four short memory tests with 10 words on each. They were told that they were to try to remember as many items as they could from each list, and then recall the words in any order at the end of presentation. Before the presentation of a list participants were asked to give an estimate of how many items they thought they would recall on each list, as a figure out of 10 (prestudy prediction). Items were presented verbally at a rate of one item every 2 s. After presentation, and before recall, patients were asked again to predict recall (poststudy prediction). Patients were encouraged to make this response quickly and told not to count up the number of items currently held in memory. In this experiment there was no digit-span distracter task, but immediate free recall was measured. At no stage was feedback given about recall performance.

RESULTS

The right panel of Figure 1 shows the mean predictions (prestudy and poststudy judgments) and recall for each list type. As expected, there were significant differences in the groups' recall, $F(1, 30) = 177.19$, $MSE = 2.17$, $p < .001$, with the OAC group outperforming the AD group. There was a main effect of difficulty, $F(1, 30) = 20.17$, $MSE = 2.36$, $p < .001$, with more words recalled from the easy lists. There was also a significant group by difficulty interaction, $F(1, 30) = 16.25$, $MSE = 2.36$, $p < .001$, the means indicating (Figure 1) that OAC recall is sensitive to difficulty but the AD group is not. There was also a main effect of relatedness, $F(1, 30) = 27.84$, $MSE = 2.27$, $p < .001$, and a group by relatedness interaction, $F(1, 30) = 13.21$, $MSE = 2.27$, $p < .001$. Participants recall significantly more words from the related lists, and the AD group did not benefit from relatedness to the same extent as

the OAC group. There was no difficulty by relatedness interaction, $F(1, 30) = 2.11$, $MSE = 1.20$, and no three-way interaction, $F < 1$.

Another factor in the design of this experiment was list order. A 2×4 (Group \times List Position) repeated measures ANOVA demonstrated that recall did not vary according to the serial order of the list. There was no main effect of list position, $F < 1$, and a nonsignificant interaction, $F < 1$. This indicates that collapsed across list type, there were no order or practice effects for recall.

To analyze the sensitivity of global predictions to list type, the poststudy predictions were analyzed. Because the objective properties of the list were known, we expected participants to predict higher recall on easier and semantically related lists. A $2 \times 2 \times 2$ (Group \times Difficulty \times Relatedness) ANOVA showed a main effect of group, indicating that groups predict different levels of performance overall, $F(1, 30) = 33.41$, $MSE = 3.65$, $p < .001$. The means (Figure 1) show that the OAC group predicts higher recall than the AD group. There was no main effect of difficulty, $F < 1$, although there was a group by difficulty interaction, $F(1, 30) = 5.89$, $MSE = 2.45$, $p < .05$, suggesting that groups do not discriminate between easy and difficult lists to the same degree. This interaction was explored with simple main effects. There was a significant effect of group for both easy and difficult lists: easy, $F(1, 30) = 36.55$, $MSE = 3.02$, $p < .001$; difficult, $F(1, 30) = 8.50$, $MSE = 3.09$, $p < .01$. However, the effect of difficulty for each group showed that the AD group did not differentiate between easy and difficult lists in their predictions, whereas there was a strong trend in the OAC group to make higher predictions of performance for the easy lists: AD, $F(1, 30) = 2.30$, $MSE = 2.45$; OAC, $F(1, 30) = 3.67$, $MSE = 2.45$, $p = .065$.

There was a main effect of relatedness, $F(1, 30) = 6.42$, $MSE = 2.05$, $p < .05$. The mean poststudy predictions indicate that groups estimate that they will recall more words from the related lists. There was no group by relatedness interaction, $F < 1$, suggesting that both groups are equally aware of the benefit of semantic relatedness to recall. The difficulty by relatedness

interaction approached significance, $F(1, 30) = 3.57$, $MSE = 1.84$, $p = .069$, but the three-way interaction was not significant, $F(1, 30) = 1.87$, $MSE = 1.84$. This analysis indicates that the AD group's poststudy predictions are not sensitive to item difficulty, but are as sensitive to semantic relatedness as those of the OAC group. We return to the issue of whether this insensitivity is appropriate in the discussion given recall performance.

We also examined the way in which participants' estimation accuracy changes both across judgment type and list order. A $2 \times 4 \times 2$ (Group \times Order \times Judgment Type) repeated measures ANOVA was carried out on the nondirectional discrepancies, calculated as for Experiment 1. (For means see Table 2.) There was a main effect of group, with the AD group being more discrepant in their predictions than the OAC group, $F(1, 30) = 4.47$, $MSE = 5.72$, $p < .05$. As with Experiment 1, there was also a main effect of judgment type, with participants being more accurate in their poststudy predictions, $F(1, 30) = 85.96$, $MSE = .85$, $p < .001$. There was also a judgment type by group interaction, $F(1, 30) = 8.47$, $MSE = .85$, $p < .01$. Simple main effects show that both groups significantly revise their estimates according to judgment type: AD, $F(1, 30) = 74.20$, $MSE = .85$, $p < .001$; OAC, $F(1, 30) = 20.24$, $MSE = .85$, $p < .001$, and as with Experiment 1, a group difference in accuracy occurred for predictions made before study, but was not significant for predictions made after it: prestudy predictions, $F(1, 30) = 7.01$, $MSE = 4.28$, $p < .05$; poststudy predictions, $F(1, 30) = 1.23$, $MSE = 2.30$.

There was no main effect of order, $F(1, 30) = 2.70$, $MSE = 3.44$, although there was a group by order interaction, $F(1, 30) = 6.03$, $MSE = 3.44$, $p < .05$, and an order by judgment type interaction, $F(1, 30) = 11.87$, $MSE = 1.06$, $p < .01$. Simple main effects show that only the AD group show increased accuracy in their predictions across list order: AD, $F(1, 30) = 3.41$, $MSE = 3.35$, $p < .05$; OAC: $F < 1$. Moreover, the groups only differ significantly in their accuracy on List 1: List 1, $F(1, 30) = 8.37$, $MSE = 4.66$, $p < .01$; List 2, $F(1, 30) = 1.14$, $MSE = 3.50$; List 3, $F(1, 30) = 1.36$, $MSE = 5.05$; List 4, $F < 1$. These analyses

suggest that the interaction between group and list order is a product of the AD group being initially inaccurate. Across Lists 1 to 4, AD participants revise their estimates to levels which are not significantly more discrepant than those of the OAC group. The OAC, in comparison, do not vary their predictions across list order. The order by judgment type interaction was also examined with simple main effects. This showed an effect of judgment type for every list: List 1, $F(2, 62) = 52.64$, $MSE = 3.86$, $p < .001$; List 2, $F(2, 62) = 56.16$, $MSE = 2.41$, $p < .001$; List 3, $F(2, 62) = 40.61$, $MSE = 3.19$, $p < .001$; List 4, $F(2, 62) = 76.28$, $MSE = 1.40$, $p < .001$, and an effect of order for both predictions made before and after study: prestudy predictions, $F(4, 124) = 64.65$, $MSE = 3.34$, $p < .001$; poststudy predictions, $F(4, 124) = 44.19$, $MSE = 2.09$, $p < .001$. This indicates that for every list, poststudy predictions are more accurate than prestudy predictions. Finally, the group by order by judgment type interaction failed to reach significance, $F(1, 30) = 1.49$, $MSE = 1.06$, indicating that there are no differences in how the groups' accuracy changes with regard to order and judgment type. Moreover, across groups, both prestudy and poststudy predictions vary significantly across list order. The means suggest that both groups get more accurate at predicting their performance.

To summarize the results of Experiment 2, the AD group is less accurate overall and, as with Experiment 1, overestimates performance. However, they show a marked benefit of having studied the words in the accuracy of their predictions, and their predictions become more accurate across Lists 1 to 4. In fact, the AD group's predictions of performance are only significantly more discrepant than those of the OAC for the first list they encounter. The group by judgment type interaction indicates that the AD group revises its predictions to a greater extent from prestudy to poststudy. We also found that the AD group was not sensitive to difficulty in the poststudy predictions that they made. These predictions were inappropriate if one considers the lists objectively: Both easy lists and related lists should be easier to recall. However, not only predictions but also memory perfor-

mance of the AD group is insensitive to item difficulty.

DISCUSSION

We were interested in three effects in this experiment: (a) a replication of the difference between pre- and poststudy accuracy for the AD group; (b) an examination of whether AD participants' global predictions of performance are sensitive to objective factors which affect recall (list difficulty and relatedness); and (c) the progression of accuracy across serially presented lists. With regard to the first point, we replicated the findings of Experiment 1. Given an opportunity to encode the list, AD patients were as accurate at predicting recall as controls. AD patients' memory monitoring may be poorly calibrated, but it does reflect underlying cognitive processes in operation during encoding.

Secondly, we expected that if AD patients could receive feedback from processes occurring during encoding, they should also be sensitive to the objective qualities of the list. However, the data suggest that the AD group was not sensitive to item difficulty between lists, whereas the control group was sensitive. Taken on face value, this suggests that AD participants cannot monitor qualities of the list as well as they can encode it. However, an important consideration is that AD recall does not vary significantly across list types. Therefore, we could conclude that the AD group was correct to be insensitive in their predictions, because the relationship between their subjective predictions and their actual recall was wholly appropriate. Their memory did not respond to objective factors (item difficulty) that operated during encoding, and their predictions reflected this.

An alternative interpretation of the accuracy results is that the apparent accuracy is caused by combined floor effects in recall and a lack of sensitivity to preexisting differences in lists. Whereas control participants have to modulate their poststudy predictions according to the nature of the list and their projected performance, it could be argued that all the AD patients have to do is become progressively more aware of

their poor recall. AD patients' recall was constant across all list types, whereas the control groups' performance varied significantly. Although there is clear evidence that AD patients are monitoring memory during encoding as shown by increased accuracy after presentation, it is not clear how this pattern is affected by floor effects. We demonstrated that AD patients are not sensitive to objective differences in item-difficulty in the global judgments they make. It is possible that this lack of awareness of the nature of the to-be-remembered information may result in poorly controlled memory performance, despite showing strong evidence of memory monitoring in the poststudy predictions they make.

In contrast, there was further evidence of intact memory monitoring in AD because the prestudy predictions became more accurate across trials, indicating that participants receive self-derived feedback from performance on trials in order to revise their estimates of performance. These results are compelling because AD recall performance does not significantly vary across list position: Participants modulate their predictions to reflect performance, rather than keeping predictions constant while improving their performance. Because accuracy improves across list order even though the lists are not identical in relatedness and difficulty, it is clear that AD predictions are based on memory monitoring. If no memory monitoring was taking place, then as well as being insensitive to the qualities of the list, AD patients would not be able to revise their estimates across list order or between their prestudy prediction and their poststudy prediction.

GENERAL DISCUSSION

Most previous research (Lopez et al., 1994; McGlynn & Kaszniak, 1991; Schacter et al., 1986) has examined estimates of list performance and concluded that memory monitoring is in deficit in AD. These studies emphasize accuracy and do not examine sensitivity to to-be-remembered lists. Moreover, they use analyses that do not consider shifts in accuracy (i.e., only

one estimate is made at one phase, on one trial) or group variability (i.e., the use of directional discrepancy as the primary dependent variable). By taking estimates of prediction before and after study, the experiments here have demonstrated that AD participants monitor memory during encoding and retrieval. Their predictions are significantly altered according to whether they have encoded the items, whether they have been tested, and whether they have previously studied a list. This finding would not have been possible using the previous methods. Experiment 2 investigated sensitivity to different types of list. It was expected that, as for items, AD participants' poststudy predictions would be sensitive to objective differences in stimuli. It was found that AD group poststudy predictions were not appropriately related to the known qualities of the list.

We argue that these experiments suggest that AD participants are monitoring their memory processes during encoding. Specifically, between an initial (prestudy) prediction of performance and a judgment made after study, they realize that they will not recall as many items as they initially thought. Such a shift in predictions must reflect attitudes to memory performance. If AD patients are monitoring memory, even in some gross manner, it is reasonable to try to improve their control of memory with behavioral interventions. Additionally, from the results of this study, it is reasonable to reject a catastrophic failure in metamemory monitoring as a contributory factor to the poor episodic memory performance of people with AD. In this way, our findings are congruent with research into item judgments in AD (Bäckman & Lipinska, 1993; Lipinska & Bäckman, 1996; Moulin et al., 2000b).

An unavoidable issue in comparing people with AD with OAC is the large difference in recall performance between groups; we need to consider the role of floor effects in these experiments. For the most part, we argue that our main analysis (the shift from prestudy to poststudy predictions) is independent of the subsequent low level of recall. We have argued elsewhere (Moulin et al., 2000b) that this sensitivity approach is appropriate for measuring memory

monitoring in memory-impaired groups. Clearly the discrepancy scores are subject to influence from floor effects, because the AD participants in many cases cannot underestimate their performance (given that many participants' recall is zero). However, there is a fundamental truth to the low predictions given poststudy; participants rightly say they will recall very little. In Experiment 1, none of the AD participants initially predicted they would recall nothing, whereas 9/16 did so after study. We believe this is best conceptualized as accurate memory monitoring following study. Participants could have predicted zero recall prior to study, but they did not.

An alternative solution to the problem of differential levels of performance would have been to have artificially manipulated recall level in one or both groups. However, under such circumstances it would be hard to interpret any differences in predictions given that different instructions would be required for each group. For example, repetition may boost recall in the AD group, but this would require the AD patients to predict how well they would do on *repeated lists*. We do not know how predictions change if participants are told that lists will repeat and thus any difference in predictions between AD (repeated study) and older adult (single study) groups would be hard to interpret.

The pattern of results in these studies is consistent with the finding that AD patients are more likely to report that they are cognitively impaired after completing a cognitive test battery (Lopez et al., 1994). Experiment 2 showed that AD patients' prestudy predictions get more accurate the more they are exposed to the presentation-test procedure. This finding has implications for deficit awareness in AD. Our findings are pertinent to Zec's (1993) theory that the lack of insight in AD patients may be due to not having an enduring record of their experience. The results of the present experiments clearly indicate that AD patients update memory predictions (awareness) through their own spontaneous feedback from performance. Participants are not told how well they have performed on a previous test, but they still become more accurate as the test progresses. The results here indicate that for a memory test, repeated exposure will enable

the AD participants to make more accurate predictions of performance. However, the present experiments cannot address whether the memory awareness that the AD group gain through repeated testing is converted into an enduring record. It has been shown that AD patients do update their expectations across lists but it is not clear whether they would still be more aware if tested again after a delay. It is highly probable that because of the memory deficit the AD participant cannot remember that 3 items out of 10 is a more realistic prediction of performance than 5 out of 10. In this sense, we would like to propose that any apparent deficit in global awareness in AD is a result of the memory deficit, not vice versa.

We expected that any initially inaccurate prediction might be the result of midpoint anchoring. In both experiments, it has been shown that the AD group make a prestudy prediction on the first list which is no different from age-matched controls. In Experiment 2, a one-way ANOVA on the prestudy prediction on List 1 showed that there were no significant differences between groups, $F < 1$. Moreover, the modal value for the first prestudy prediction is 5 for both the AD and OAC groups. This analysis supports Connor et al.'s (1997) hypothesis, and extends it to a memory-impaired sample. Therefore, a possible reason why AD patients seem less accurate than controls is because the OAC group perform better on the task, and both groups predict that they will recall about half of the list. Because participants cannot make their first prestudy judgment based on the to-be-remembered items in this experiment, it is reasonable that they make a general prediction based on a "rule of thumb." The level of this first prestudy prediction seems independent of whether the person is demented or not because there is no significant difference between AD and OAC participants. If there is evidence for a deficit in memory monitoring in AD it is that they are not aware of how bad their memory is until being prompted by a test phase, and being informed about their memory deficit during the encoding process. Therefore, we argue that throughout life, people have learned that 50% is a useful estimate of perfor-

mance, and this estimate is not updated when the person gets dementia.

The prestudy 50% estimates in the AD group may also indicate that the AD group is setting a target of performance, rather than making a prediction. There is evidence of this goal setting behavior in older adults. For example, Connor et al. (1997) found 3 of their older adults (with relatively low recall levels) had initially predicted 60 out of 60 items, which they suggest is an example of goal-setting behavior, instead of a proper estimate.

In essence, the experiments presented here suggest that AD patients monitor their memory during encoding. The apparent inaccuracy in their prestudy predictions is consistent with the same midpoint anchoring seen in the control group. Once this erroneous prestudy judgment has been made, people with AD revise their estimates – and the basis of this revision is proficient memory monitoring.

REFERENCES

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edition, rev.). Washington, DC: Author.
- Bäckman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease, *Neuropsychologia*, *31*, 335-345.
- Battig, W.F., & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, *80*, 1-46.
- Bruce, P.R., Coyne, A., & Botwinick, J. (1982). Adult age differences in metamemory. *Journal of Gerontology*, *37*, 354-357.
- Connor, L.T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*, 50-71.
- Correa, D.D., Graves, R.E., & Costa, L. (1996). Awareness of memory deficit in Alzheimer's disease patients and memory impaired older adults. *Aging, Neuropsychology, and Cognition*, *3*, 215-228.
- Devolder, P.A., Brigham, M.C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. *Psychology and Aging*, *5*, 291-303.

- Feher, E.P., Larrabee, G.J., Sudilovsky, A., & Crook, T. (1994). Memory self-report in Alzheimer's disease and in age-associated memory impairment. *Journal of Geriatric Psychiatry and Neurology*, *7*, 58-65.
- Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-Mental State: A practical method for grading the cognitive state of the patient for the clinician. *Journal of Psychiatric Research*, *12*, 189-198.
- Goldblum, M., Gomez, C., Dalla Barba, G., Boller, F., Deweer, B., Hahn, V., & Dubois, B. (1998). The influence of semantic and perceptual encoding on recognition memory in Alzheimer's disease. *Neuropsychologia*, *36*, 717-729.
- Hachinski, V.C., Linnette, D.I., Zilhka, E., DuBoulay, G.H., McAllister, V.L., Marshall, J., Russel, R.W.R., & Symon, L. (1975). Cerebral bloodflow in dementia. *Archives of Neurology*, *32*, 632-637.
- Herlitz, A., & Viitanen, M. (1991). Semantic organisation and verbal episodic memory in patients with mild and moderate Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *13*, 559-574.
- Hertzog, C., Saylor, L.L., Fleece, A.M., & Dixon, R.A. (1994). Metamemory and aging: Relations between predicted, actual and perceived memory task performance. *Aging and Cognition*, *1*, 203-237.
- Light, L.L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, *42*, 333-376.
- Lipinska, B., & Bäckman, 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.D., & DeKosky, S.T. (1994). Awareness of cognitive deficits and anosognosia in probable Alzheimer's disease. *European Journal of Neurology*, *34*, 277-282.
- McDonald-Miszczak, L., Hunter, M.A., & Hultsch, D.F. (1994). Adult age differences in predicting memory performance: The effects of normative information and task experience. *Canadian Journal of Experimental Psychology*, *48*, 95-118.
- McGlynn, S.M., & Kaszniak, A.W. (1991). When metacognition fails: Impaired awareness of deficit in Alzheimer's disease. *Journal of Cognitive Neuroscience*, *3*, 183-189.
- 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.
- Moulin, C.J.A., Perfect, T.J., & Jones, R.W. (2000a). The effects of repetition on allocation of study time and judgements of learning in Alzheimer's disease. *Neuropsychologia*, *38*: 750-758.
- Moulin, C.J.A., Perfect, T.J., & Jones, R.W. (2000b). Evidence for intact memory monitoring in Alzheimer's disease: Metamemory sensitivity at encoding. *Neuropsychologia*, *38*: 1242-1250.
- Nebes, R.D. (1989). Semantic memory in Alzheimer's disease. *Psychological Bulletin*, *106*, 377-394.
- Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-173). San Diego, CA: Academic Press.
- Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology*, *14*, 330-345.
- Rebok, G.W., & Balcerak, L.J. (1989). Memory self-efficacy and performance differences in young and old adults: The effect of mnemonic training. *Developmental Psychology*, *25*, 714-721.
- Rubin, D.C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronounceability for 925 nouns. *Memory & Cognition*, *14*, 79-94.
- Schacter, D.L., McLachlan, D.R., Moscovitch, M., & Tulving, E. (1986). Monitoring of recall performance by memory disorders patients. *Journal of Clinical and Experimental Neuropsychology*, *8*, 130.
- Shimamura, A.P., & Squire, L.R. (1986). Memory and metamemory: A study of feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 452-460.
- Zec, R.F. (1993). Neuropsychological functioning in Alzheimer's disease. In R.W. Parks, R.F. Zec, & R.S. Wilson (Eds.), *Neuropsychology of Alzheimer's disease and other dementias* (pp. 3-80). Oxford: Oxford University Press.

APPENDIX

Stimuli used in Experiments 1 and 2.**Experiment 1**

List 1: mountain, doctor, dove, friend, student, concept, origin, occasion, position, impropriety.

List 2: caravan, brassiere, sky, teacher, tree, elaboration, disclosure, permission, typhoon, hint.

Experiment 2

E-U: boy, death, elephant, grandmother, inn, joy, lake, university, policeman, shoes (mean probability of free recall; Rubin & Friendly, 1986; 0.75).

D-U: causality, figment, sulphur, formation, joviality, necessity, outsider, pledge, shame, chance (mean probability of free recall; Rubin & Friendly, 1986; 0.25).

E-R: hour, minute, second, year, day, century, month, decade, week, millisecond.

D-R: indigo, maroon, tan, lavender, beige, aqua, magenta, olive, rose, mauve.