PERGAMON

Neuropsychologia 38 (2000) 1242–1250

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Evidence for intact memory monitoring in Alzheimer's disease: metamemory sensitivity at encoding

Chris J.A. Moulin[a, c,*], Timothy J. Perfect[b], Roy W. Jones[c]

[a]*Department of Psychology, University of Reading, PO Box 238, Earley Gate, Whiteknights, Reading RG6 6AL, UK*
[b]*Department of Psychology, University of Plymouth, Plymouth, UK*
[c]*Research Institute for the Care of the Elderly, St Martin's Hospital, Bath, UK*

Received 3 August 1999; received in revised form 1 February 2000; accepted 8 February 2000

Abstract

Previous research claiming that there is a metamemory deficit in Alzheimer's Disease (AD) has been based on paradigms in which metamemory judgements are compared with performance. These methods confound predictive accuracy with very poor memory performance. In the experiments presented here this confound is removed by focusing on the sensitivity of metamemory judgements to item differences at encoding, rather than on predictive accuracy. In Experiment 1 participants studied words of high or low recallability, and either made judgements of learning (JOLs) or declared recall readiness. It was found that the AD group discriminate between items in their metamemory judgements to the same extent as age matched controls. Both groups rated the highly recallable words as being more likely to be recalled, and allocated more study time to low recallability items. In Experiment 2 participants were asked to rank the likelihood of recall of items that varied in objective recallability. Once again, AD patients were as sensitive to objective differences in stimuli as controls. Therefore, using measures based on sensitivity to item differences, we find no evidence of a metamemory deficit at encoding in AD. The findings are discussed in terms of metamemory functioning in AD, and its relationship with memory performance. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Dementia; Judgements of Learning (JOLs); Metacognition

1. Introduction

Recent research suggests that an encoding deficit, rather than increased forgetting, underlies the poor performance on episodic memory tasks in Alzheimer's disease (AD) [3,6,7]. It may therefore be important in Alzheimer's disease to assess factors such as metamemory that operate during encoding. This paper examines the idea that a deficit in metamemory functioning could be an exacerbating factor in the poor episodic performance in AD.

One cause of the encoding deficit might be that people with AD fail to control and monitor memory during encoding. Several studies have examined metamemory functioning in AD [2,11,19]. All these studies are limited in their ability to address metamemory function at encoding because they examine metamemory judgements made after encoding and before retrieval (the feeling of knowing (FOK) procedure). Moreover, the first two studies use general knowledge materials rather than assessing performance on a memory task that includes an encoding phase. These general knowledge studies suggest that there is no deficit in metacognition, with the AD group being as accurate as controls in assigning confidence to recalled answers and predicting future recognition performance.

The study by Pappas et al. [19] is more directly relevant to the present work because they studied metamemory judgements in episodic memory with recall

---

* Corresponding author. Tel.: +44-118-931-6699; fax: +44-118-931-6715.

E-mail address: c.j.a.moulin@reading.ac.uk (C.J.A. Moulin).

PII: S0028-3932(00)00037-3

and recognition tasks. For the recall task, they were unable to draw conclusions about the predictive accuracy of metamemory judgements because of floor effects in the recall of the AD group. However, for the recognition task, with performance off floor, they found that AD patients do not predict recognition as accurately as controls. They conclude, "It should be determined if predictions of recall performance are intact in AD patients. If so, this would indicate a dissociation between metamemory for recall and recognition" (p. 163).

There are profound logical difficulties in concluding that people with AD have impaired metamemory on this kind of evidence. It is inferred that metamemory is inaccurate when participants' predictions of performance fail to relate to how they actually perform: a word that is judged to be highly recallable should be more likely to be recalled than a word rated as less likely to be recalled. Problems arise with this approach when testing participants who have an episodic memory impairment, because their likelihood of remembering *any* item is at floor. Memory performance is so poor that it precludes statistical comparisons of groups' metacognitive abilities.

More importantly, the reason that metamemory judgements lack predictive power at test may be because of processes that occur after encoding. Participants may make appropriate predictions of recall during study that would be predictive were it not for the separate episodic memory deficit. That is, participants may accurately monitor the difficulty of different items to be learned, and may take appropriate steps to control their encoding to achieve learning. However, using accuracy-based measures of metamemory does not allow one to focus on what occurred at encoding.

In the present work we offer a new approach which overcomes the confound with memory performance, and focuses instead on processes occurring at encoding. Our reasoning is straightforward; if metacognition is intact at encoding in AD, then memory monitoring and control by participants with AD should be as sensitive as normals to item differences. To distinguish our approach from previous work examining metacognitive accuracy, we classify the measures we use as metacognitive sensitivity measures.[1]

In the experiments reported here, participants study words that have been selected on the basis of objective measures of difficulty. Participants either make predictions of future performance for these words in the conventional manner, or study them until they believe the items have been learned. If AD patients are monitoring memory as they encode items they should be sensitive to the differences between items, even if these predictions are not ultimately accurate. For example, they should rate objectively difficult words as being harder, and dedicate more study time to them to recall, even if they fail to recall any of the items.

In the first experiment sensitivity of memory monitoring is measured by asking participants to predict, at learning, future recall performance for a set of words with objectively known properties. This kind of metamemory prediction has been extensively used before, and is known as a judgement of learning or JOL (e.g. [13,18]). Participants were asked to rate words selected to be either difficult-to-recall or easy-to-recall, based on Rubin and Friendly's [20] free recall norms. By examining whether judgements reflect these pre-existing differences it is possible to find evidence of metamemory monitoring without having the confound of floor effects in AD participant's recall.

The first study also examines memory control at encoding using allocation of study time as a dependent variable. This measure, also known as recall readiness, is a well-established measure (e.g. [4,17]). In this paradigm participants are allowed to study words as long as they feel necessary to ensure subsequent recall before declaring recall readiness. Again, the difference in study times for objectively different items can be measured. It is predicted that if memory control is intact and correctly informed by monitoring, then participants should spend longer studying the objectively difficult items.

Whilst there is previous research addressing the role of memory monitoring in AD, there has been less work in metamemory control processes. Our own research [15] suggests that with multiple presentations during encoding, AD participants, like controls, allocate less time to items they have already encoded. However, the present work is the first to examine whether appropriate allocation of study time is found for different kinds of to-be-learned words in a single-trial test.

2. Experiment 1

2.1. Method

2.1.1. Participants

Sixteen AD patients and 16 age-matched older adult controls (OAC) were tested. Diagnosis of AD was made by a clinician using information from neuropsychological examination, mini-mental state examination (MMSE) [5], family interview, laboratory screening

---

[1] Hertzog and Dixon [9] argue that metacognitive judgements about memory fall into three different conceptual categories — declarative knowledge about memory processing, awareness of one's on-line memory processing, beliefs about one's own memory system. However, at present, there is no clear consensus about how these concepts inter-relate. Our measure of metacognitive sensitivity clearly relates most closely to the second of these constructs.

Table 1
Participant details for all experiments: mean (M) and standard deviation age in years, mini mental state examination score, and years of education[a]

	Age		MMSE		Yrs education	
	M	SD	M	SD	M	SD
Experiment 1						
AD	75.8	5.6	21.8	3.1	9.9	1.8
OAC	70.2	4.4	–	–	11.0	2.8
Experiment 2						
AD	75.2	9.1	15.1	4.8	11.2	2.5
OAC	74.9	2.3	–	–	11.3	2.8

[a] Notes: AD = Alzheimer's disease ($n = 16$), OAC = older adult controls ($n = 16$). MMSE = mini-mental state examination [5]. Yrs education = years spent in formal education.

(i.e. haematology; B12 and folate levels; renal, liver and thyroid function; calcium and syphilis serology) and medical examination. Patients were diagnosed as being demented with the DSM III-R [1] criteria and as having AD by the NINCDS-ADRDA criteria [14]. 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 or if they had a Hachinski score [8] that indicated they might have a vascular component to their dementia. The OAC group was also screened for dementia before being admitted to the study. This group consisted either of carers 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 ($n = 10$). The participant characteristics are shown in the top panel of Table 1. A one-way ANOVA showed that there was no significant difference in the education level (years education) of these two groups [$F(1, 30) = 1.62$]. The principal analyses were run with age as a covariate in the analyses of variance (ANCOVA) because there were significant age differences between the two groups (the AD group was significantly older than the OAC group; [$F(1, 30) = 9.62$, $p < 0.01$]). The effect of age failed to account for any of the differences between groups for every dependent variable.

2.1.2. Stimuli/materials

Five easy and five difficult words were selected for each task from Rubin and Friendly's [20] recall norms.[2] The words were presented on a computer screen centrally in black on a white ground, in letters 2 cm high. Participants were given an opportunity to check the readability of the materials before the trial commenced and no participant reported any difficulty. To avoid errors in use of the computer keyboard, participants' responses for all tasks were made verbally and were keyed into the computer by the experimenter.

2.1.3. Design and procedure

This experiment consisted of two separate tasks, a judgement of learning (JOL) task and a recall readiness task. The same participants carried out both. There was a mixed design, with two groups (AD and OAC) and two levels of objective difficulty of word (easy vs difficult).

Participants were tested individually. Every participant completed the recall readiness task before the JOL task. The procedures used in each task were the same except that in one case participants studied each word until they declared recall readiness and in the other presentation time was fixed and participants had to give each word a JOL.

2.1.3.1. Recall readiness task. Participants were familiarised with the computer and screen. Instructions were presented on the computer screen and also read aloud by the experimenter. Participants were instructed that they would see words to be studied and that these words would remain on the screen until the participant verbally declared 'recall readiness'. Participants were instructed to study the word until they felt they had maximised their chance of remembering the word in the subsequent test, with the instruction to "go at your own speed, but don't spend longer on a word than you need to". Declaration of recall readiness prompted the next trial. Prior to the test, participants were told that there would be an immediate free recall test at the end of the list, with an unlimited time to recall. The to-be-remembered items were presented individually in a random order. After presentation of the 10 words, there was a short (four digits) digit span task. Participants then attempted verbal free recall of the items that they had studied.

2.1.3.2. JOL task. This was the same as the recall readiness task, except that each item appeared for only 2 s and was then masked. At this point, participants were prompted to rate how likely it was that they would be able to recall the word later, using a 5 point scale ("I'd like you to tell me how likely you are to remember this word: 1 = won't remember it, 5 = definitely will remember it"). Participants were instructed to concentrate equally on rating the word and remembering it.

---

[2] The words used were: recall readiness task, easy: *mountain, doctor, dove, friend, student*; difficult: *concept, origin, occasion, position, impropriety*. JOL task, easy: *caravan, brassiere, sky, teacher, tree*; difficult: *elaboration, disclosure, permission, typhoon, hint*.

2.2. Results

Three aspects of the data were analysed for each task: recall performance, sensitivity of metacognitive judgements to item differences and accuracy of metacognitive judgements in predicting recall.

2.2.1. Recall performance

The mean number of easy and difficult items recalled for both tasks is shown in Table 2. The pattern is typical of the poor recall found in AD, with patients on average recalling less than one word. A $2 \times 2 \times 2$ ANOVA (group × task × item) was carried out on this data. There were main effects of group [$F(1, 30) = 69.72$, $p < 0.001$] and item [$F(1, 30) = 14.78$, $p = 0.001$] in line with expectations, but no effect of task [$F(1, 30) < 1$]. The lack of an effect of task suggests that giving participants as long as they like to maximise their recall performance does not improve recall, compared to a fixed 2-s presentation time for each word that in practice is a shorter study time.

The task × group interaction approached significance [$F(1, 30) = 3.179$, $p = 0.085$]. There was a significant group × ease interaction [$F(1, 30) = 11.97$, $p < 0.01$]. Simple main effects show that whilst the recall of the AD group is not sensitive to item differences, [$F(1, 30) < 1$, n.s.], the older adults' recall is (OAC: $F(1, 30) = 24.25$, $p < 0.001$). There was no trial × ease interaction [$F(1, 30) = 2.39$, n.s.], and the three-way interaction was not significant [$F(1, 30) < 1$]. In summary, AD recall is much lower than controls and possibly because of floor effects fails to show differences across items. The older adults recall significantly more easy words than difficult words.

2.2.2. Metamemory sensitivity

To examine memory monitoring we looked at whether the JOLs and study times for words were in the direction predicted by the objective differences between words. Lower JOLs, and longer study times were expected for the objectively difficult words. Mean JOLs and study times for each word type are shown in Table 3.

A 2×2 (group × item) ANOVA was carried out on the mean JOLs given to the different word types. There was no main effect of group [$F(1, 30) = 2.58$, n.s.]. There was a main effect of item [$F(1, 30) = 38.88$, $p < 0.001$] such that the JOLs were in line with objective difficulty. The group × item interaction was not significant [$F(1, 30) = 2.58$, n.s.]: the AD group gave JOLs that were as sensitive to objective difficulty as controls.

Recall readiness times were analysed in the same manner. There was a main effect of group [$F(1, 30) = 15.56$, $p < 0.001$] and the means show that the AD group spends a lot longer studying the words than the controls. The main effect of item was not significant [$F(1, 30) = 1.72$, n.s.], such that participants do not spend significantly longer studying the more difficult words. There was no significant interaction between group and item [$F(1, 30) < 1$], indicating that study time is equally insensitive to objective difficulty in both groups. Examining metamemory sensitivity to item differences, using both JOL and recall readiness measures, reveals no evidence of a deficit in the AD group relative to non-diseased controls.

2.2.3. Metamemory accuracy

This more conventional analysis compares metacognitive predictions with subsequent recall. Accuracy was determined by examining the mean JOL and mean study time for items subsequently recalled compared to the same for items subsequently forgotten. Because six AD patients' recall was zero in the JOL task, their data is not included in this analysis. The mean JOLs

Table 2
Experiment 1: mean (and SD) number of items recalled (out of five) by each group, for each level of item difficulty and encoding task[a]

Encoding task	Item difficulty			
	Easy		Hard	
	M	SD	M	SD
Judgement of learning task				
AD	0.88	1.15	0.50	0.73
OAC	2.87	1.02	1.56	0.89
Recall readiness task				
AD	0.25	0.68	0.50	0.73
OAC	2.94	1.00	1.87	1.20

[a] Notes: AD = Alzheimer's disease group (n = 16), OAC = older adult controls (n = 16).

Table 3
Experiment 1: mean judgement of learning (on 5 point scale) and time to declare recall readiness (in seconds) for each kind of item[a]

Encoding task	Item difficulty			
	Easy		Difficult	
	M	SD	M	SD
Judgement of learning rating (out of five)				
AD	4.04	0.75	2.83	0.60
OAC	4.15	0.74	3.45	0.98
Time (in seconds) to declare recall readiness				
AD	8.0	5.0	9.6	7.7
OAC	2.7	0.8	3.2	1.2

[a] Notes: AD = Alzheimer's disease group (n = 16), OAC = older adult controls (n = 16).

and times to declare recall readiness are shown in Table 4.

For mean JOLs split by recall status, a 2 × 2 (group × recall status) ANOVA was carried out. There was no main effect of group [$F(1, 24) = 1.44$, n.s.]. There was a main effect of recall status (words recalled vs words not recalled) [$F(1, 24) = 9.91$, $p < 0.01$] with higher JOLs for words subsequently recalled. The interaction approached significance [$F(1, 24) = 3.12$, $p = 0.09$], with a suggestion that the AD group were less discriminating in their JOL ratings. For recall readiness analysed in the same manner, there was a main effect of group [$F(1, 20) = 15.94$, $p = 0.001$], but no main effect of recall status [$F(1, 20) < 1$] and no interaction [$F(1, 20) < 1$]. This suggests no clear benefit of extra study time in the recall of items in either group. However, there are three important considerations. First, the analysis lacks power, since the AD group in the recall readiness condition is reduced to six because of zero recall. Second, the relationship between item difficulty, study time and recall is unclear. The effect of the first two on the last is difficult to resolve, since difficulty should increase study time, study time should increase recall, and yet, greater difficulty should *reduce* recall. Third, it is already known that a large increase in study time yields small, if not negligible, increases in memory performance (the '*labor-in-vain*' effect [17]).

Finally, we report the results from a conventional measure of accuracy, the gamma statistic [16]. This is a non-parametric correlation that measures the association between a participant's JOL and their actual recall. This statistic, calculated for each participant, produces a figure closer to one for a more accurate predictor of performance. There was no statistically significant difference between groups [$F(1, 23) = 2.51$, n.s.] although the means (and standard deviations) suggest that the AD group is less accurate than the older adults are: AD, 0.14 (0.71); OAC, 0.54 (0.55). Because of zero recall, the AD group was reduced to 10.

Table 4
Experiment 1: mean (and standard deviations) judgement of learning ratings (out of five) and time to declare recall readiness (in seconds) given to words subsequently recalled or not

	Recall status			
	Recalled		Not recalled	
	M	SD	M	SD
Judgement of learning rating				
AD[a]	3.59	0.90	3.32	0.81
OAC	4.27	0.59	3.30	1.01
Recall readiness (study time, seconds)				
AD[b]	11.1	10.0	11.2	7.8
OAC	3.0	1.0	2.9	1.0

[a] $n = 10$ because 6 AD patients recalled zero items.

[b] $n = 6$ because 10 AD patients recalled zero items. AD = Alzheimer's disease, OAC = older adult controls.

2.3. Discussion

This experiment used a novel approach to researching metamemory at encoding. By emphasising metamemory sensitivity rather than metamemory accuracy it was possible to overcome confounds with criterion memory performance, and examine metacognitive functioning at encoding. For memory monitoring, AD patients were as sensitive as older adults in the ratings (JOLs) they gave to items. For the measure of memory control, recall readiness, Experiment 1 showed no significant differences between groups, but the huge variations in study times meant that it was difficult to assess whether all participants were appropriate in their study times. Because neither group recalls more words in the recall readiness condition than in the JOL task, it seems that participants do not use their increase in study time to boost recall. This is in keeping with the 'labor in vain' effect [17] whereby large increases in study time yield very little differences in recall.

The fact that AD participants seem metacognitively sensitive during encoding is contrary to the accuracy results found previously for episodic stimuli. Pappas et al. [19] found that for predicting recognition, AD patients were significantly less accurate than controls. In the present study, the accuracy measures are inconclusive because of floor effects, but if anything tend to support Pappas et al., in that the means suggest that the AD group are less accurate than controls. Therefore, predictions of recall in AD are as impaired as predictions of recognition. However, the sensitivity measures are not in keeping with these accuracy results: the AD group makes appropriate judgements during encoding. Aside of the fact that these results are not confounded by memory performance, there are two further methodological differences between this study and Pappas et al.'s work. First, in their study, participants predict recognition, whereas here participants predict recall. Second, in their study participants predict recognition for items that have not been recalled successfully. It is possible that attempting retrieval somehow prompts participants to make a more accurate prediction of subsequent performance. However, it must be noted that any such effect would run contrary to the benefits of greater variability across items that comes with examining metacognition for all items rather than merely those which have not been recalled.

One problem with the present experiment is that the AD patients might be as sensitive as controls in their

JOLs just because participants are responding to very marked differences between items. Stimuli in this experiment came from two very distinct groups of words. This exaggerated difference between words might have strongly driven participants' JOLs. To overcome this, Experiment 2 looks at monitoring of items using a ranking measure for a set of words across a range of difficulty.

3. Experiment 2

In Experiment 1 it was found that AD patients were as sensitive as controls to objective differences between items. One potential problem with this finding was that the to-be-remembered items were split into two very distinct sets of words (e.g. *teacher* compared to *impropriety*). In this experiment a range of items is presented and a ranking procedure used to judge the recallability of words. An alternative to JOLs as a metacognitive judgement is ranking judgements (for a review see Nelson and Narens [18]). These have been used with differing success, according to whether it is absolute or relative accuracy that is being investigated. Nelson and Narens ([18], p. 143) suggest that rankings may be useful when "... an investigator is not interested in the absolute aspects of FOK ratings ...". In this experiment, rankings are used because the relative sensitivity of the AD patients is the focus of study. By asking participants to rank recallability and then comparing this with the objective rank (based on norms of recall performance), it is possible to measure sensitivity by calculating the association between subjective and objective measures of recallability. This gives us an impression of the sensitivity of the participant's predictions, rather than just comparing mean JOLs for distinct words.

If the results of Experiment 1 are indicative of intact sensitivity to items at encoding, then it is expected that the ranking of words' recallability made by the AD patients will be as accurate as control's ranking. By comparing correlations between objective and subjective ratings of the words, it is possible to examine degree of sensitivity. If the AD group is not capable of the same fine-grained metacognitive judgements as controls, then the level of association between subjective and objective rankings should differ between groups.

3.1. Method

3.1.1. Participants

A different set of participants from Experiment 1 was selected using the same criteria. As before, there was an AD group ($n = 16$) and an older adult control group (OAC; $n = 16$) recruited from a panel of volunteers at the University. The OAC group was paid to take part in the study, and completed this task as part of a larger battery of tests. The characteristics of the participants are shown in the bottom panel of Table 1.

In order to remove floor effects in recall, participants were excluded from this experiment if they did not recall at least one item from one of the lists. This resulted in the rejection of data from two AD participants. Nevertheless, there is a large difference between the total recall scores for the two groups [$F(1, 30) = 11.09$, $p < 0.001$]. However, in this experiment there are no significant differences in the groups' mean age [$F(1, 30) < 1$] or education level [$F(1, 30) < 1$].

3.1.2. Materials/stimuli

Ten words[3] were chosen from the Rubin and Friendly [20] recall norms, and the words were graded in terms of mean recallability (FRA) in steps of 0.06–0.08, from words that were highly recallable (e.g. mother) to less easily recalled words (e.g. figment). Words were presented to participants on cards, with words printed in lower case approximately 1.5 cm high. All words were checked for readability by participants, who reported no problems.

3.1.3. Design and procedure

There were two successive trials of the same test. On each of these trials, participants were asked to rank the words in a different order. On one occasion they ranked from most easy to remember to most difficult, and on the other trial, this order was reversed. This factor was counterbalanced, so different participants either ranked easy word first or difficult words first on the first trial. List differences were analysed to eliminate the possibility that words ranked first or last, regardless of difficulty, may be differentially recalled.

Participants were tested individually in a quiet room. They were presented with the words on a table in front of them, arranged randomly in an array so that they could see all 10 words at once. The AD participants were instructed to read each word out loud once. In the easy-word-first list, participants were asked to choose the word that was the easiest word for them to remember from the set of 10. The instruction given was to "pick out the word that you think is the easiest for you personally to remember". The experimenter then removed the word from the array and placed it aside, instructing the participant to select the easiest word for them to remember from the remaining set. In this way, the ranking procedure was always based on the direct comparison of words, rather than relying on memory for what had been presented. After

---

[3] The words used were (from easiest to most difficult): mother, sea, lad, scarlet, abdomen, engine, speech, opinion, truce, figment.

each word was selected, the experimenter put it into an ordered list in front of the participant, with the first item picked out furthest from the participant and the last word picked out nearest them. After all 10 words had been ranked, the participant was asked to check the 10 words to decide to see whether they would like to change the order. In the case of the OAC group, once the participant had indicated that they were satisfied that this was the correct order, the words were removed from view and immediate free recall was measured. Based on the differences in study time shown in Experiment 1, the AD patients were given a further minute in which they were instructed to learn the list. After this period, the AD participants attempted verbal free recall. All participants were instructed that they could recall the words in any order.

3.2. Results

3.2.1. Recall performance

Recall was not affected by whether items were ranked from difficult to easy words or vice versa, nor did this factor interact with any other. To examine memory performance across the two trials, a 2 × 2 (group × trial) repeated measures ANOVA compared groups' recall performance for trial one and trial two (see means in text below). There was a main effect of group [$F(1, 30) = 111.09$, $p < 0.001$] showing that, as expected, the OAC group recalled significantly more items. There was also a main effect of trial [$F(1, 30) = 16.94$, $p < 0.001$] indicating that learning took place from trial to trial. There was no interaction [$F(1, 30) = 2.13$, n.s.] suggesting that both groups add equal numbers of items to their recall in the second trial. This result has to be interpreted with caution as the OAC group is close to ceiling on trial 2. Whilst the AD group recall increases from 2.56 (SD 1.93) to 3.19 (SD 1.72) across trials, the equivalent figures for the OAC groups are 7.69 (SD 1.66) and 9.00 (SD 0.97).

3.2.2. Metamemory sensitivity and accuracy

Metamemory sensitivity was analysed by calculating a Spearman's rho correlation coefficient for each participant between the rank they gave each item, and the mean FRA taken from Rubin and Friendly's [20] norms. A positive correlation indicates that participants are judging the items in line with normative difficulty. The correlation coefficient, therefore, is a measure of how well the participant's ranking of words relates to objective or 'population-wide' performance. For the AD group the mean correlations for trial 1 and trial 2 were 0.52 (SD 0.20) and 0.55 (SD 0.23), whilst the equivalent figures for the OAC group were 0.59 (SD 0.19) and 0.44 (SD 0.40). A 2 × 2 (group × trial) repeated measures ANOVA revealed that there were no significant group [$F(1, 30) < 1$] or trial [$F(1, 30) < 1$] differences and the interaction failed to reach significance [$F(1, 30) = 2.28$, n.s.]. From the positive correlations and the fact that all groups were higher than zero by a significant margin (Student's t-tests, all $t(16) > 4$, $p < 0.001$) it can be concluded that participants are sensitive to the objective differences between words, and make their ranking judgements accordingly. The AD are as sensitive as the OAC group in ranking the words, and neither group is more or less sensitive in trial 1 compared to trial 2. This method has therefore demonstrated that AD participants are as sensitive as controls even when making judgements about a range of items.

The accuracy analysis considers whether participants are more likely to recall the five words they rank as being easy than the five they rank as difficult. The respective means (and standard deviations) for easy and difficult words were, trial 1: OAC: 4.06 (1.06), 3.62 (0.80); AD: 1.50 (1.41), 1.06 (1.06); trial 2: OAC: 4.50 (0.73), 4.50 (0.52); AD: 1.75 (1.34), 1.44 (0.96). These data suggest that participants recall more of the words they rank as easy than the words they rank as difficult. However, the formal analysis was not significant because of the floor (AD) and ceiling (OAC) effects.

3.3. Discussion

This experiment has found further evidence for item sensitivity in metamemory for AD patients. The mean Spearman's rho correlations between Rubin and Friendly [20] recall proportion and participants' individual ranking of items are high, and these correlations do not differ significantly between groups. This is compelling evidence for intact memory monitoring in AD because in addition to the sensitivity to easy and difficult items in Experiment 1, the difference between items is not marked in this experiment — the to-be-remembered items represent a range of difficulties. Thus AD patients' success in Experiment 1 is not due to the basic ability to distinguish between easy and difficult items, but is equally demonstrated for finer discriminations across items.

4. General discussion

These two experiments clearly indicate that there is no metacognitive deficit at encoding in Alzheimer's disease, as measured by our sensitivity measures. In Experiment 1 it was found that AD patients allocate study time as appropriately as controls and rate items appropriately (JOLs). Ranking measures were likewise appropriate for AD patients in Experiment 2 when a range of items were used.

We would not have come to the same conclusion if we had used the traditional measure of metacognitive accuracy to assess metacognitive function in AD. For compelling evidence from neuropsychological studies of metacognitive accuracy, one has to turn to work such as Shimamura and Squire [21]. They found a deficit in predictive accuracy for a Korsakoff's group relative to another memory impaired group (non-Korsakoff's amnesia), despite equally poor memory in the two groups. Even so, a deficit in metamemory accuracy in this case does not illuminate our understanding of encoding processes in memory impaired groups.

For research into metacognition in memory-impaired groups, these experiments provide a clear starting point for examining encoding processes. It has been necessary to fractionate the metamemory process into its component parts. We have argued that one of the roles of metamemory monitoring is to be sensitive at encoding to factors that affect subsequent recall: here we have shown that AD patients are sensitive to the objective difficulty of to-be-remembered words, even though their performance is impaired. Therefore, we find no basis for a metamemory deficit at encoding in AD.

There is some support for memory monitoring not being impaired in AD. Studies that examine memory monitoring for general knowledge stimuli, also find intact metamemory in AD [2,10]. We propose that this could be due to the fact that the memory impairment in AD affects episodic material more than it does general knowledge material [19], and in turn, poor recall performance affects accuracy measures. If one is analytical about what it is that metacognition actually does, then the control and monitoring framework [18] suggests that a sensitivity paradigm is going to be insightful about the processes that operate during encoding. The work presented here suggests AD patients control and monitor memory proficiently during encoding.

There are possible shortcomings of the sensitivity approach. The first is that despite being an appropriate indicator of encoding behaviour, the sensitivity measures do not necessarily tap participants' awareness of their memory system. Participants might be sensitive to the objective differences because of the physical characteristics of the words rather than their registration in memory. For example, Lovelace [12] suggests that awareness can be based on normative or idiosyncratic components. Normative components include item difficulty — the surface characteristics of an item that are available before study (e.g. familiarity, pronouncability, word length). These contrast with idiosyncratic or privileged access components of metamemory which are based on the participant's own awareness of their memory system. These two contrasting means of predicting performance may capture the differences between the AD and control groups' predictions. The AD group may be sensitive but not accurate because they are relying on normative characteristics and not idiosyncratic characteristics. This could explain why their JOLs relate to norms of recall performance, but not to actual performance.

A second shortcoming of the sensitivity approach is that participants might not be considering future performance in their JOLs. To explore this we can compare the difference in results for ease-of-learning (EOL) and judgements of learning (JOL) measures, two measures reported in the metacognition literature. EOLs are defined as metamemory judgements made before the memory item has been mastered (i.e. during presentation and before prolonged study), rather than the JOLs made after the item has been presented (as reported here). Nelson and Narens [18] suggest that because EOL judgements occur prior to encoding, they cannot tap memory processes but can only tap item difficulty. It is hypothesised that in contrast, JOLs, made after an item has been learned, can tap metaknowledge about how well that item is stored in memory. JOLs are a much more accurate predictor of subsequent performance than EOLs. For example, Leonesio and Nelson [10] found that subsequent recall is predicted much better by JOLs ($\gamma = 0.31$) than by EOLs ($\gamma = 0.12$). Also, they found that EOLs are a relatively poor measure of monitoring performance, in that EOLs did not correlate ($\gamma = -0.22$) with the number of trials required to learn the various items in a constant-study-time situation. It is possible that the sensitivity approach emphasises the surface characteristics of the to-be-remembered items, and participants make predictions akin to an EOL rather than a JOL. Further research could address this possibility. The differences in surface characteristics could be removed and instead participants could make sensitivity judgements based on other factors that increase recall. For example, as well as work examining sensitivity to repeats [15] we could look at manipulations of level of processing. This way, we could see if metamemory judgements were sensitive to factors that affect recall that are not related to the surface features of the to-be-remembered items.

To summarise, these experiments find that memory monitoring is not impaired in AD, if one concentrates on comparing monitoring to objective differences in words (the sensitivity approach). There is some uncertainty about whether these predictions accurately relate to recall. However, AD patients' control and monitoring are appropriate, which suggests that a gross unawareness of difficulty is not behind the encoding deficit in AD. There is a possibility that participants make their metamemory judgements based on the surface characteristics of words, but this pattern is no different for either group. In any case, this surface characteristic emphasis would be a reasonable heuristic to use, given that in most cases the characteristics of the word influence recall. In conclusion, we propose that the AD group are inaccurate not because they do not differentiate between items in their allocation of resources at encoding, but because their recall does not live up to expectations. That is, whereas AD patients monitoring judgements are sensitive, their recall is not. The implication of this research for the understanding of Alzheimer's disease is that the memory impairment in AD — particularly the encoding deficit — is not exacerbated by a metacognitive deficit.

Acknowledgements

The authors would like to acknowledge the Economic and Social Research Council for the Studentship that supported the first author. We also wish to thank the staff and patients of the BRACE (Bristol Research into Alzheimer's disease and Care of the Elderly, Blackberry Hill Hospital, Bristol) and RICE (Research Institute for the Care of the Elderly, St Martin's Hospital, Bath) memory clinics.

References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3rd ed. Washington, DC: American Psychiatric Association, 1987.

[2] Bäckman L, Lipinska B. Monitoring of general knowledge: evidence for preservation in early Alzheimer's disease. Neuropsychologia 1993;31:335–45.

[3] Christensen H, Kopelman MD, Stanhope N, Lorentz L, Owen P. Rates of forgetting in Alzheimer dementia. Neuropsychologia 1998;36:547–57.

[4] Dunlosky J, Connor LT. Age differences in the allocation of study time account for age differences in memory performance. Memory and Cognition 1997;25:691–700.

[5] Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: a practical method for grading the cognitive state of the patient for the clinician. Journal of Psychiatric Research 1975;12:189–98.

[6] Granholm E, Butters N. Associative encoding and retrieval in Alzheimer's and Huntington's disease. Brain and Cognition 1988;7:335–47.

[7] Greene JDW, Baddeley AD, Hodges JR. Analysis of the episodic deficit in early Alzheimer's disease: evidence from the doors and people test. Neuropsychologia 1996;34:527–51.

[8] Hachinski VC, Linnette DI, Zilhka E, DuBoulay GH, McAllister VL, Marshall J, et al. Cerebral bloodflow in dementia. Archives of Neurology 1975;32:632–7.

[9] Hertzog C, Dixon RA. Metacognitive development in adulthood and old age. In: Metcalfe J, Shimamura AP, editors. Metacognition: knowing about knowing. Cambridge, USA: MIT Press, 1994. p. 227–51.

[10] Leonesio RJ, Nelson TO. Do different metamemory judgements tap the same underlying aspects of memory? Journal of Experimental Psychology: Learning, Memory, and Cognition 1990;16:464–70.

[11] Lipinska B, Bäckman L. Feeling of knowing in fact retrieval: further evidence for preservation in early Alzheimer's disease. Journal of the International Neuropsychological Society 1996;2:350–8.

[12] Lovelace EA. Metamemory: monitoring future recallability during study. Journal of Experimental Psychology: Learning, Memory and Cognition 1984;10:756–66.

[13] Mazzoni G, Nelson TO. Judgements of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. Journal of Experimental Psychology: Learning, Memory, and Cognition 1995;21:1263–74.

[14] McKhann G, Dracman D, Folstein M, Katzman R, Price D, Stadlan EM. 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 1984;34:939–44.

[15] Moulin CJA, Perfect TJ, Jones RW. The effects of repetition on allocation of study time and judgements of learning in Alzheimer's disease. Neuropsychologia 2000;38:748–756.

[16] Nelson TO. A comparison of current measures of feeling-of-knowing predictions. Psychological Bulletin 1984;95:109–33.

[17] Nelson TO, Leonesio RJ. Allocation of self-paced study and the "Labor in Vain Effect". Journal of Experimental Psychology: Learning, Memory and Cognition 1988;14:676–86.

[18] Nelson TO, Narens L. Metamemory: a theoretical framework and some new findings. In: Bower GH, editor. The psychology of learning and motivation, vol. 26. San Diego, USA: Academic Press, 1990. p. 125–73.

[19] Pappas BA, Sunderland T, Weingartner HM, Vitiello B, Martinson H, Putnam K. Alzheimer's disease and feeling of knowing for knowledge and episodic memory. Journal of Gerontology: Psychological Sciences 1992;47:159–64.

[20] Rubin DC, Friendly M. Predicting which words get recalled: measures of free recall, availability, goodness, emotionality, and pronouncability for 925 nouns. Memory & Cognition 1986;14:79–94.

[21] Shimamura AP, Squire LR. Memory and metamemory: a study of feeling-of-knowing phenomenon in amnesic patients. Journal of Experimental Psychology: Learning, Memory and Cognition 1986;12:452–60.