



Alzheimer's disease and memory-monitoring impairment: Alzheimer's patients show a monitoring deficit that is greater than their accuracy deficit[☆]

Chad S. Dodson^{a,*}, Maggie Spaniol^{b,c,d}, Maureen K. O'Connor^{b,c,d}, Rebecca G. Deason^{b,c,d}, Brandon A. Ally^{e,f,g}, Andrew E. Budson^{b,c,d}

^a Department of Psychology, P.O. Box 400400, 102 Gilmer Hall, University of Virginia, Charlottesville, VA 22904-4400, United States

^b Center for Translational Cognitive Neuroscience, Geriatric Research Education Clinical Center, Bedford VA Hospital, Bedford, MA, United States

^c VA Boston Healthcare System, Boston, MA, United States

^d Boston University Alzheimer's Disease Center, Boston University School of Medicine, Boston, MA, United States

^e Department of Neurology, Vanderbilt University, Nashville, TN, United States

^f Department of Psychology, Vanderbilt University, Nashville, TN, United States

^g Department of Psychiatry, Vanderbilt University, Nashville, TN, United States

ARTICLE INFO

Article history:

Received 17 January 2011

Received in revised form 10 May 2011

Accepted 12 May 2011

Available online 19 May 2011

Keywords:

Alzheimer's disease
Recognition memory
Source memory
Memory monitoring
Calibration
Confidence
Anosognosia

ABSTRACT

We assessed the ability of two groups of patients with mild Alzheimer's disease (AD) and two groups of older adults to monitor the likely accuracy of recognition judgments and source identification judgments about who spoke something earlier. Alzheimer's patients showed worse performance on both memory judgments and were less able to monitor with confidence ratings the likely accuracy of both kinds of memory judgments, as compared to a group of older adults who experienced the identical study and test conditions. Critically, however, when memory performance was made comparable between the AD patients and the older adults (e.g., by giving AD patients extra exposures to the study materials), AD patients were still greatly impaired at monitoring the likely accuracy of their recognition and source judgments. This result indicates that the monitoring impairment in AD patients is actually worse than their memory impairment, as otherwise there would have been no differences between the two groups in monitoring performance when there were no differences in accuracy. We discuss the brain correlates of this memory-monitoring deficit and also propose a Remembrance–Evaluation model of memory-monitoring.

© 2011 Elsevier Ltd. All rights reserved.

Much research has documented the costs to memory from Alzheimer's disease (AD). Individuals with AD are less likely than age- and education-matched controls to recognize or recall previously encountered events (Budson, Wolk, Chong, & Waring, 2006). AD patients also show difficulty remembering source information – specific information about the exact circumstances under which an event was encountered (e.g., Johnson, Hashtroudi, & Lindsay, 1993). They are less able than controls to remember: (a) whether an item was previously seen or imagined (e.g., Dalla Barba, Nedjam, & Dubois, 1999), (b) whether a sentence was completed by oneself or by another (e.g., Multhaup & Balota, 1997), (c) whether words had been previously presented in a red or green color

(Tendolkar et al., 1999), (d) whether a fact was said by a man or a woman (Mitchell, Sullivan, Schacter, & Budson, 2006), (e) whether a word was presented during one task or another (Pierce, Sullivan, Schacter, & Budson, 2005; Pierce, Waring, Schacter, & Budson, 2008), (f) whether an item was presented during the study or test session (Budson, Dodson, Daffner, & Schacter, 2005), and (g) source information outside of the laboratory (Budson et al., 2004, 2007; see Souchay & Moulin, 2009, for a review).

Although the memory deficit from AD is increasingly clear, what is not clear is the extent to which AD patients are aware of this memory deficit. In other words, how well can AD patients monitor and judge the likely accuracy of memories? To foreshadow our results, this paper shows that AD patients are strikingly unaware of the accuracy of their memories. Even when their recognition performance is comparable to that of an older control group, AD patients are unable to distinguish between correct and incorrect recognition responses.

Reliably monitoring one's memory is critically important for knowing how much to trust a particular memory. One problem in particular that patients with AD experience is a high percent-

[☆] This research was supported by National Science Foundation grant 0925145 (CSD) and National Institute on Aging grants R01 AG025815 (AEB) and P30 AG13846 (AEB). This material is also the result of work supported with resources and the use of facilities at the Bedford VA Hospital in Bedford, MA.

* Corresponding author.

E-mail address: cdodson@virginia.edu (C.S. Dodson).

age of false memories, both in the laboratory (Budson, Daffner, Desikan, & Schacter, 2000) and in the real world (Budson et al., 2007). For instance, patients who falsely remember that they have already turned off the stove or have taken their medications when they have not will no longer be able to live independently. However, accurately monitoring memories is a potential mechanism for counteracting the influence of false memories when, for example, individuals are aware that a particular memory is likely to be false. Thus, understanding the extent to which patients with AD can monitor and judge the accuracy of their memories is an important clinical issue in addition to being of great theoretical interest.

The small literature that has examined memory-monitoring on the part of Alzheimer's patients has produced conflicting findings (see Pannu & Kaszniak, 2005; Souchay, 2007 for reviews). On the one hand, patients with mild AD appear no different from healthy controls in using confidence ratings to judge the likely accuracy of answers to general knowledge questions (e.g., Who wrote *Alice in Wonderland*?) that assess well-learned or semantic information (Backman & Lipinska, 1993; Lipinska & Backman, 1996; see also Cosentino, Metcalfe, Butterfield, & Stern, 2007). Similarly, Moulin, James, Perfect, and Jones (2003) observed no differences between AD participants and healthy older controls in monitoring the accuracy of recognition judgments. On the other hand, Souchay, Isingrini, and Gil (2002) showed that AD patients were impaired in monitoring episodic memories. When completing a cued-recall test of recently learned material, AD patients were much less accurate than older controls in providing feeling-of-knowing (FOK) judgments about the likelihood of recognizing an unrecalable item (Souchay et al., 2002). To our knowledge, no one has examined AD patients' ability to monitor the likely accuracy of source judgments.

In general, this pattern of AD patients – when compared to healthy older adults – showing a preserved ability to monitor the accuracy of semantic memories but an impaired ability to monitor the accuracy of episodic memories parallels a similar pattern that exists when healthy older adults are compared to younger adults. Older adults are either no different from or better than younger adults in monitoring the accuracy of responses to general knowledge questions – either with confidence ratings or by providing FOK judgments (e.g., Allen-Burge & Storandt, 2000; Butterfield, Nelson, & Peck, 1988; Dodson, Bawa, & Krueger, 2007; Perlmutter, 1978; Pliske & Mutter, 1996). However, older adults are much worse than younger adults in monitoring the accuracy of recently learned (episodic) material: (a) they provide less accurate FOK judgments on cued-recall tests (Souchay, Isingrini, & Espagnet, 2000); and (b) their confidence ratings are less well calibrated on both cued-recall tests and source memory tests because of the occurrence of high-confidence errors (Dodson, Bawa, & Krueger, 2007; Dodson, Bawa, & Slotnick, 2007; see also Dodson & Krueger, 2006). Dodson, Bawa, and Krueger (2007) and Dodson, Bawa, and Slotnick (2007) proposed a misrecollection account to explain the pattern of older adults' monitoring performance. They suggested that advancing age is associated with an increased susceptibility of miscombining features of different events that in turn produce confidently held misrecollections – memory problems that are likely caused by deterioration of the hippocampus and surrounding tissue (e.g., Shimamura & Wickens, 2009; Squire, Stark, & Clark, 2004; and see Raz et al., 2005 for evidence of accelerated shrinkage of the hippocampus with healthy aging). This account predicts age-related differences in monitoring performance – *because of the occurrence of high confidence errors* – on all tests that require memory for specific details about recently learned past events, such as source memory tests.

The goals of the present study are twofold. First, given the similarity between older adults and AD patients in that both show relatively preserved monitoring of semantic memories and impaired monitoring of episodic memories (when compared to

their respective control groups), we sought to test the misrecollection account as an explanation of the monitoring performance on the part of AD patients. Specifically, if AD patients show a more severe version of the same misrecollection impairment that is exhibited by older adults then we should observe that on a source memory test AD patients are more prone than older adults to make high confidence errors.

The second goal is to examine whether the memory-monitoring deficit in AD patients is a byproduct of their overall weaker memory performance or whether this memory-monitoring deficit is a byproduct of a more specific, malfunctioning memory-monitoring mechanism. The answer to this question is unknown because in all extant studies AD patients show worse memory performance than their healthy older counterparts. Thus, worse memory-monitoring by AD patients may be a byproduct of their worse accuracy (see Hertzog, Dunlosky, & Sinclair, 2010 for a similar argument about differences in memory-monitoring between older and younger adults). No study has sought to compare the memory-monitoring abilities of AD patients and healthy older adults when both groups show comparable memory performance.

To address these goals, we used a combination recognition and source identification paradigm that involves collecting confidence ratings for *both* memory judgments (e.g., Dodson, Bawa, & Krueger, 2007; Simons, Peers, Mazuz, Berryhill, & Olson, 2010). At encoding, participants heard sentences that were presented by either a female or a male. At test, participants were presented with sentences and made an initial old–new recognition judgment (i.e., Did you encounter it before or is it new?) and for items receiving a judgment of old, participants made a source judgment about who presented the item earlier (i.e., male or female?). For both the recognition and the source judgments, participants rated the likely accuracy of their responses on a confidence scale from 50 (guessing) to 100 (certain). According to our misrecollection account, we should observe that AD patients make more high confidence errors than healthy older adults on the source judgment. Critically, if the monitoring impairment on the part of AD patients is not merely a byproduct of their worse accuracy then we should observe that even when AD patients and healthy older adults show comparable source identification accuracy, the AD patients still will exhibit a monitoring impairment. In other words, matching the AD patients and the older adults on source memory accuracy avoids a potential confound and so any observed differences in monitoring performance on this judgment cannot be attributed to differences in source accuracy.

1. Methods

1.1. Participants

The participants consisted of twenty-four clinically diagnosed, mild AD patients (age range from 56 to 86) who were assigned to either the AD group or the AD-m group (i.e., 12 in each group). Twenty-four healthy older adults (age range from 62 to 90) were assigned to either the Older group or the Older-m group (i.e., 12 in each group). Each group consisted of 6 females and 6 males, except in the AD-m group where there were 5 females and 7 males. The AD group and the older group experienced the identical study and test conditions. By contrast, the AD-m group received extra repetitions of the encoding material so as to boost their memory performance to that of the older adult group. Likewise, the Older-m group received a longer study list that served to lower and match their memory performance to that of the AD group. We assessed overall cognitive ability with the following neuropsychological tests: the Mini-Mental State Examination (MMSE), the Consortium to Establish a Registry for Alzheimer's disease (CERAD) word list memory test, the Boston Naming Test-Short Form 15 (BNT-15), Word Fluency (FAS/CAT), Trail Making Test-A (TMT-A), and Trail Making Test-B (TMT-B).¹ Inclusion in either

¹ These data were acquired from the AD patients and the healthy older adults within the preceding 6 months and 12 months, respectively, of completing this study. Apart from scores for age, education and the MMSE, the remaining neuropsychological scores are missing for 1 patient in the AD-m group.

Table 1
Demographic and neuropsychological test means (*M*), standard deviations(*SD*), and *F* tests of significance for individuals in the AD, AD-m, Older and Older-m groups.

Measure and group	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>
Age (years)				
AD	77.00	8.83	<1.0	<i>ns</i>
AD-m	79.83	8.18	1.16	<i>ns</i>
Older	77.50	5.35		
Older-m	76.50	6.95	<1.0	<i>ns</i>
Education (years)				
AD	15.00	2.59	<1.0	<i>ns</i>
AD-m	14.33	3.89	5.43	0.029
Older	16.04	3.03		
Older-m	17.25	1.91	1.36	<i>ns</i>
MMSE				
AD	23.92	2.71	6.56	0.018
AD-m	26.42	2.02	20.58	<0.0005
Older	29.83	0.39		
Older-m	29.33	0.98	2.68	<i>ns</i>
CERAD encoding				
AD	9.50	3.03	5.75	0.026
AD-m	12.82	3.60	43.68	<0.0005
Older	24.08	4.12		
Older-m	22.50	3.42	1.05	<i>ns</i>
CERAD delayed recall				
AD	1.00	1.04	2.71	<i>ns</i>
AD-m	2.09	2.02	61.07	<0.0005
Older	7.42	2.07		
Older-m	8.00	1.60	<1.0	<i>ns</i>
CERAD delayed recognition				
AD	5.42	2.97	<1.0	<i>ns</i>
AD-m	6.09	3.48	15.23	<0.0005
Older	10.00	0		
Older-m	10.00	0	<1.0	<i>ns</i>
Letter fluency				
AD	33.42	16.21	<1.0	<i>ns</i>
AD-m	38.73	12.93	11.49	<0.005
Older	45.25	12.3		
Older-m	57.00	12.90	5.22	0.032
Category fluency				
AD	27.17	11.22	<1.0	<i>ns</i>
AD-m	31.36	11.16	16.54	<0.0005
Older	47.33	9.94		
Older-m	50.08	10.91	<1.0	<i>ns</i>
TMT-A (s)				
AD	64.25	32.18	<1.0	<i>ns</i>
AD-m	54.73	23.36	11.01	<0.005
Older	31.33	8.40		
Older-m	31.25	7.24	<1.0	<i>ns</i>
TMT-B (s)				
AD	210.58	84.70	8.11	0.01
AD-m	120.27	65.02	7.04	0.015
Older	82.67	25.13		
Older-m	69.17	14.89	2.56	<i>ns</i>
BNT-15				
AD	12.67	3.52	<1.0	<i>ns</i>
AD-m	12.00	3.35	5.46	0.029
Older	14.42	0.79		
Older-m	14.42	1.24	<1.0	<i>ns</i>

Note: The top *F* and *p* values are for comparisons between the two groups of AD patients, the middle values are for comparisons between the AD-m group and the Older-m group and the bottom values are for comparisons between the two groups of older adults. MMSE = Mini-Mental State Examination; CERAD = Consortium to Establish a Registry for Alzheimer's disease word list memory test; TMT-A = Trail Making Test-A; TMT-B = Trail Making Test-B; BNT-15 = Boston Naming Test-Short Form 15. Apart from scores for Age, Education and the MMSE, the remaining scores are missing for 1 participant in the AD-m group (i.e., $n = 11$).

older adult group required that each participant score above 27 out of 30 on the MMSE. Individual neuropsychological testing scores were required to be within two standard deviations of the published mean for mild AD and controls (Morris et al., 1989).

Table 1 shows the demographic information, neuropsychological test scores and the statistical comparisons between (a) the two groups of AD patients (the top *F* and *p* values within each measure), (b) the AD-m group and the Older-m group (the middle *F* and *p* values), and (c) the two groups of older adults (the bottom *F* and *p* values). Analyses of variance (ANOVA) showed no significant group differences in age, $F_s(1, 22) < 1.2$, but the AD-m group completed fewer years of education than did

their Older-m counterparts, $F(1, 22) = 5.43$, $p < 0.05$. As expected, the AD-m patients performed significantly worse than their Older-m counterparts on the MMSE, CERAD (e.g. encoding, delayed recall and delayed recognition), TMT-A, TMT-B, letter fluency and category fluency. The two groups of AD patients were no different on all measures, except the AD-m patients performed significantly better than the AD patients on the MMSE, CERAD-Encoding and the TMT-B. Thus, overall, the AD-m group may be less cognitively impaired than the AD group. We note in the discussion that this potential difference between our two AD groups actually strengthens our conclusions about memory-monitoring. The two different groups of older adults were no different on all of the measures with the exception of letter fluency.

Finally, all individuals in both groups of older adults reported: (a) no history of neurological or any other psychological impairment, (b) no history of stroke or focal brain injury, (c) no history of drug or alcohol abuse, (d) no first-degree relatives with memory disorders, (e) an adequate ability to see and hear, and (f) being a native speaker of English. All AD patients were: (a) diagnosed with mild AD by a neurologist; (b) had an adequate ability to see and hear; and (c) were a native speaker of English. Each participant provided informed consent and was reimbursed \$10.00 for every hour of participation.

1.2. Materials

We used the same materials as in Dodson, Bawa, and Krueger (2007). The trivia statements and voice recordings of the statements were presented to participants with E-Prime software (Psychology Software Tools Inc., www.pstnet.com).

1.3. Procedure

All individuals were told at the encoding phase that they would hear statements that were presented by a male (i.e., Bill) speaker and a female (i.e., Kim) speaker. They were shown pictures of Bill and Kim and also were informed that they would see the visual form of the statement. Everyone was told to pay attention to each statement as well as to who was reading it because they would be tested on this information at a later point.

The individuals in the older adult and AD groups encountered 30 statements, half spoken in a male voice (i.e., Bill) and half spoken in a female voice (i.e., Kim). For the AD-m group, however, this encoding phase was repeated so that each statement was encountered a total of three times – always presented by the same speaker each time. By contrast, the individuals in the Older-m group encountered 100 statements (equal amounts presented by the male and female voices) so as to lower their source identification performance to that of the AD patients. However, the make-up of these 100 statements consisted of the same 30 statements that were presented to the older and AD groups as well as an additional 70 “filler” statements that were randomly intermixed with these 30 target statements. We only scored performance on the 30 target statements that were identical for all groups. The encoding list for all groups also included an additional six statements – three at the beginning and at the end of the encoding list – that served to control for primacy and recency effects. Each statement was presented visually on a laptop screen for 7 s and during this time participants heard the statement spoken by either the male or female speaker and also saw the corresponding picture of the statement's speaker (i.e., either “Bill” or “Kim”).

The test phase for all groups immediately followed the completion of the encoding phase and consisted of sixty statements (30 encoding statements and 30 new statements). Equal numbers of the encoding statements had been spoken by “Bill” and “Kim” during the study phase (i.e., 15 each). Participants were told that on the memory test they would see statements that would appear only visually. Participants were instructed that they would first make an old–new recognition judgment and would then provide a confidence rating about the likely accuracy of this recognition judgment by selecting one of six numbers: 50 (guessing), 60, 70, 80, 90, or 100 (certain). The different confidence ratings were fully explained to all individuals so that it was understood that the ratings reflected gradations of certainty from no certainty (i.e., guessing) to absolute certainty that the response is correct. Judgments of new automatically initiated the next test statement. By contrast, when participants recognized a statement as old, they were then asked to indicate whether Bill or Kim presented this statement during the encoding phase. Participants then provided a confidence rating with the same six-point scale about the likely accuracy of this source judgment. Following this confidence rating about the source judgment, the next test statement would automatically appear on the screen. For both recognition and source judgments, participants were encouraged to use the entire scale of confidence ratings. The test statements were presented in a random order with the constraint that no more than two of the same sources were presented consecutively. In addition, there were three practice statements at the beginning of the test phase that were used to demonstrate the task and to ensure that each participant understood the instructions.

Lastly, each participant first completed a demographic questionnaire before completing the memory paradigm. All neuropsychological tests were conducted after the completion of the memory paradigm and only to those individuals whose files needed updating (5 older adults and 4 AD patients underwent neuropsychological tests).

2. Results

2.1. Recognition performance and recognition monitoring

We examined recognition performance with hit rates (i.e., proportion of studied items judged “old”), corrected recognition rates (i.e., hit rates minus false alarm rates to novel items) and d' scores (e.g., Snodgrass & Corwin, 1988). With respect to monitoring performance, there are two fundamentally different ways of measuring how individuals use confidence ratings to monitor and judge the accuracy of their responses (e.g., Koriat & Goldsmith, 1996; Pannu & Kaszniak, 2005; Yaniv, Yates, & Smith, 1991). Relative measures of monitoring – often referred to as resolution – assess how well individuals can use confidence ratings to distinguish between correct and incorrect responses by for example assigning higher confidence ratings to all correct responses and lower confidence ratings to all incorrect responses. This is a relative process of monitoring performance because the absolute levels of the confidence ratings are not as important as the relative differences in the confidence ratings for correct and incorrect judgments (e.g., the same resolution score occurs when individuals give correct and incorrect judgments confidence ratings of either 100 and 50 or just 60 and 50). We measured this relative process of monitoring in two different ways: (1) by using the Somers' d correlation between confidence and accuracy; and (2) by examining the overall average confidence rating that is assigned to correct and incorrect responses (e.g., Pannu & Kaszniak, 2005). By contrast, absolute measures of monitoring – often referred to as calibration – assess how well the absolute values of an individual's confidence ratings align with performance. For instance, excellent calibration between accuracy and confidence is shown when individuals exhibit perfect, moderate and chance performance for responses that are assigned, respectively, high (e.g., 100), medium (e.g., 80) and low (e.g., 50) confidence ratings. In this way, the absolute value of the confidence ratings corresponds to the absolute level of performance for items that are assigned these ratings. Calibration error scores are derived for each individual by taking the absolute difference between predicted accuracy, as indicated by the confidence rating, and actual accuracy at each level of confidence (see Koriat & Goldsmith, 1996). This difference is then further weighted by the frequency of responses at each level of confidence (see Brewer & Wells, 2006; Koriat & Goldsmith, 1996; Pannu & Kaszniak, 2005, for discussion about computing relative and absolute measures of monitoring).

Table 2 shows measures of recognition accuracy: hit rates, corrected recognition scores and d' scores. False alarm rates were low for all groups – as seen by comparing the hit rates with the corrected recognition rates – and were nearly non-existent for the two older groups. Although the AD group showed a substantial recognition deficit, performance by the AD-m individuals was comparable to that of both groups of older adults. Analyses of variance (ANOVA) showed significant group differences in hit rates, corrected recognition rates, and d' scores, all $F_s(3, 44) > 5.12$, all $p_s < 0.01$. With all measures, post hoc Fisher's PLSD tests confirmed that the AD group

Table 2
Recognition and source identification accuracy scores.

Group	Hit rates	Corrected recognition	Recognition d'	Source score	
				Correct	Incorrect
AD	0.65 (0.06)	0.49 (0.08)	1.51 (0.24)	0.55 (0.02)	
AD-m	0.83 (0.07)	0.77 (0.07)	2.87 (0.32)	0.60 (0.05)	
Older	0.89 (0.03)	0.88 (0.03)	3.30 (0.14)	0.76 (0.03)	
Older-m	0.89 (0.03)	0.88 (0.03)	3.37 (0.21)	0.60 (0.03)	

Note: Parentheses contain the standard error of the mean.

The AD and the Older adults received the identical study and test conditions. The AD-m group received multiple repetitions of the study material and the Older-m group received a longer (i.e., more items) encoding phase.

Table 3
Recognition and source identification monitoring scores.

Group	Recognition		Source	
	Somers' d	Calibration	Somers' d	Calibration
AD	0.05 (0.09)	0.28 (0.07)	0.01 (0.06)	0.24 (0.03)
AD-m	0.02 (0.09)	0.12 (0.05)	0.08 (0.07)	0.23 (0.03)
Older	0.72 (0.11)	0.09 (0.01)	0.23 (0.03)	0.14 (0.02)
Older-m	0.53 (0.10)	0.10 (0.03)	0.30 (0.08)	0.24 (0.03)

Note: Parentheses contain the standard error of the mean.

The AD group and the Older adults received the identical study and test conditions. The AD-m group received multiple repetitions of the study material and the Older-m received a longer (i.e., more items) encoding phase.

performed significantly worse than the AD-m group and worse than both groups of older adults, all $p_s < 0.05$. By contrast, with all measures there were no significant differences between the AD-m group and both groups of older adults.

Table 3 shows recognition Somers' d scores, which measure how well both groups of older adults and both groups of AD patients can assess the accuracy of recognition responses with confidence ratings. Perfect performance is represented by a Somers' d score of 1 and means that all correct responses received a higher confidence rating than all incorrect responses. By contrast, a score of 0 means there is no relationship between confidence ratings and the correctness of a response. As is clear from Table 3, both groups of AD patients showed a severe monitoring impairment. An ANOVA of these recognition Somers' d scores revealed a significant effect of group, $F(3, 30) = 11.83$, $MSE = 0.08$, $p < 0.001$, with both the AD group and the AD-m patients showing significantly lower Somers' d scores than either group of older adults, $p_s < 0.001$. In fact, the Somers' d correlation for either group of AD patients is no different from 0, which means that AD patients show no correlation between their confidence ratings and recognition accuracy. It is particularly striking that the AD-m group showed an inability to distinguish between correct and incorrect recognition responses even though their objective memory performance is good (i.e., hit rate = 83%) and comparable to that of either of the healthy older groups.

Another way to appreciate the difference in Somers' d scores between the groups is by examining the overall average confidence rating that is assigned to correct (hits) and incorrect (misses) recognition responses. Table 4 shows that both groups of older adults exhibited a pattern that was consistent with optimal monitoring. They were more confident about their correct than their incorrect recognition judgments, both $t_s > 3.00$, $p_s < 0.05$. By contrast, both groups of AD patients were just as confident in their incorrect recognition judgments as they were in their correct recognition judgments, as the average confidence rating was essentially identical for both correct and incorrect recognition responses, both $t_s < 1$, $p > 0.05$.

Table 4
Average confidence rating assigned to correct and incorrect recognition and source judgments.

Group	Recognition		Source attributions	
	Hits	Misses	Correct	Incorrect
AD	87 (2.7)	87 (3.5)	74 (4.6)	74 (4.8)
AD-m	88 (3.9)	87 (4.7)	76 (3.9)	74 (4.2)
Older	97 (1.2)	84 (4.5)	77 (3.6)	63 (2.8)
Older-m	96 (1.3)	82 (4.3)	80 (3.3)	72 (3.9)

Note: Confidence scale ranges from 50 to 100 and parentheses contain the standard error of the mean.

The AD group and the Older adults received the identical study and test conditions. The AD-m group received multiple repetitions of the study material and the Older-m received a longer (i.e., more items) encoding phase.

With respect to the calibration of confidence ratings to recognition performance, Table 3 shows recognition calibration error scores, which measure the alignment of confidence ratings to hit rates. For instance, perfect alignment would receive a calibration error score of 0 and would occur when someone, for example, shows a hit rate of 100% for those items that received a confidence rating of 100, whereas greater deviations between confidence and accuracy would receive larger calibration error scores. An ANOVA of these recognition calibration error scores revealed an effect of group, $F(3, 44) = 4.02$, $MSE = 0.02$, $p < 0.05$. Post hoc Fisher's PLSD tests showed that the AD group performed significantly worse than any of the other groups, all $ps < 0.05$. By contrast, there were no significant differences between the AD-m group and both groups of older adults. Thus, the AD-m group appears to have some preserved ability to calibrate confidence ratings with recognition accuracy.

Lastly, is it possible that the observed differences in monitoring are caused by systematic differences in how the AD patients and the older adults used the confidence rating scale (e.g., are AD patients more polarized when making confidence ratings by primarily using the end points of the scale)? We addressed this question by comparing the AD-m group and the Older-m group on the distribution of confidence ratings to studied items because these were the two groups that showed comparable recognition performance (i.e., similar hit rates, corrected recognition and d' scores) and yet showed very different capacities to monitor the accuracy of these responses – at least in terms of Somers' d scores and average confidence ratings. Both groups distributed their confidence ratings across the scale in a similar manner, as there were no significant differences between the groups in the overall use of each confidence rating (i.e., 50, 60, 70, 80, 90, 100), all $t's(22) < 1.77$. Thus, differences in the use of the confidence rating scale cannot account for the pattern of monitoring differences in Somers' d scores and the overall average confidence ratings between these groups.

Overall, both groups of AD patients showed a substantial impairment in monitoring resolution. Even though the recognition accuracy of the AD-m group is comparable to that of the healthy older adults, these AD individuals nevertheless exhibited a profound inability to distinguish between correct and incorrect responses with confidence ratings, as shown by nearly equivalent overall average confidence ratings to hits and misses and by Somers' d scores that are close to zero. Moreover, even with the substantial improvement in recognition accuracy on the part of the AD-m group as compared to the AD group, there was essentially no change in monitoring resolution. However, it is important to note that the AD-m group did show improved calibration error scores as compared to the AD group, and so there may be some preservation of this different dimension of monitoring.

2.2. Source identification and monitoring performance

Source identification performance was assessed with conditional source scores, which refer to the proportion of items correctly recognized as old (i.e., hit rate) that are attributed to the correct source (e.g., Dodson, Bawa, & Krueger, 2007). Table 2 shows the conditional source scores (i.e., remembering who-said-what) for each group. An ANOVA revealed significant group differences, $F(3, 44) = 5.65$, $MSE = 0.018$, $p < 0.01$, with the individuals in the Older adult group showing significantly higher source scores than individuals in either the Older-m group, the AD-m group or the AD group, all $ps < 0.01$, as determined with Fisher's PLSD post hoc tests. There were no significant differences between the Older-m group and either of the AD groups.

The right side of Table 3 shows source Somers' d correlations between the confidence ratings that were assigned to correct and incorrect source attributions. If an individual's ability to monitor and judge the likely accuracy of a response is determined by overall

source memory performance, then we should observe a similar pattern of monitoring performance by both groups of AD patients and by the Older-m group, as they all show comparable source scores. An ANOVA of the Somers' d correlations for the confidence ratings that were assigned to correct and incorrect source attributions yielded a significant effect of group, $F(3, 43) = 6.33$, $MSE = 0.075$, $p < 0.01$. Fisher's PLSD tests showed that both groups of AD patients exhibited worse Somers' d scores than did the Older group, $ps < 0.01$, and worse scores than the Older-m group, although this difference was marginally significant on the part of the AD-m group (i.e., $p = 0.06$). Neither group of AD patients nor the older groups were significantly different from each other.

The foregoing pattern of differences in Somers' d correlations is also seen in the overall average confidence ratings that are assigned to correct source attributions and incorrect source attributions. As is clear from the right side of Table 4, both groups of older adults showed higher average confidence ratings for correct than incorrect source judgments, both $ts > 3.40$, both $ps < 0.01$. By contrast, both groups of AD patients showed a fundamental inability to distinguish between correct and incorrect source attributions. They were comparably confident in the accuracy of both correct and incorrect source judgments, both $ts < 1$, both $ps > 0.05$.

However, Table 4 clearly shows that the overall average confidence ratings that are assigned to the source judgments are lower than the confidence ratings that are assigned to recognition judgments. In fact, for all groups, including both groups of AD patients, the overall confidence rating for the recognition judgment was significantly higher than this rating for the source judgment, all $t's(11) > 3.31$, all $ps < 0.01$. Given that their recognition performance was much better than their source identification performance, this difference in overall confidence rating suggests that the AD patients are aware that the source judgment is a more difficult task. Therefore, these data show a dissociation in AD patients between the relative lack of awareness of the correctness of a particular response on the one hand and on the other hand the awareness at the level of the task that the recognition responses are more often correct than are the source judgments.

With respect to the source calibration error scores, an ANOVA revealed a significant effect of group, $F(3, 44) = 2.79$, $MSE = 0.01$, $p = 0.05$, with older adults (0.14) showing significantly better calibration scores than any of the other groups, $ps < 0.05$. As seen in the right side of Table 3, both groups of AD patients were no different from the Older-m group. Although Older-m adults and both groups of AD patients showed similar overall calibration scores, they do so for very different reasons.

Fig. 1 shows source scores that are computed from responses that received either low (50, 60), medium (70, 80) or high (90, 100) confidence ratings. It is important to note that because participants did not use each of the confidence ratings equally frequently (i.e., some confidence ratings were used more than others) it is not possible to look at Fig. 1 and derive the overall source score for a group by averaging the particular source scores that are based on the responses given low, medium and high confidence ratings. The key thing to notice in this figure is that the source scores for both groups of older adults increased with increasing confidence. For both groups of AD patients, however, there was a relatively unchanging level of performance across the different confidence ratings.

As seen in Fig. 1, the Older-m group's source scores showed a linear function in that their performance steadily increased with increasing confidence – indicating that they were comparably miscalibrated at roughly all points of the confidence scale. By contrast, both groups of AD patients become increasingly worse calibrated with increasing confidence. Specifically, as shown in Fig. 1, when either group of AD patients used the high end of the confidence rating scale (i.e., ratings of 90, 100) they were significantly less accu-

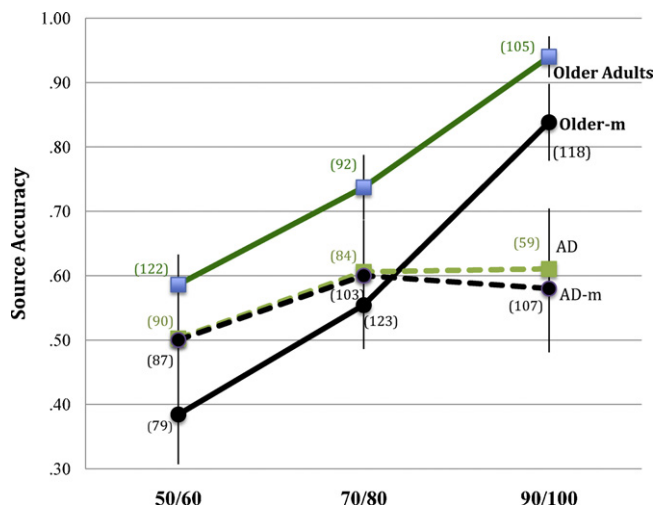


Fig. 1. Source scores at the different confidence intervals for the Older, Older-m, AD, and the AD-m groups. *Note:* Error bars are standard error of the mean. Parentheses contain the total number of responses, aggregated across individuals in that group. The AD and the Older adults received the identical study and test conditions. The AD-m group received multiple repetitions of the study material and the Older-m group received a longer (i.e., more items) encoding phase.

rate when making source attributions than were the Older-m group – their accuracy-matched older adult counterparts, $t(19)=2.32$, $p<0.05$ for the AD-m group and $t(16)=2.13$, $p<0.05$ for the AD group. And, there were no significant differences between either of the AD groups and the Older-m group when using the low (50, 60) or medium (70, 80) confidence ratings, $t_s < 1.22$. Thus, both groups of AD patients were particularly vulnerable to making high confidence source identification errors, which contributed to their impaired ability to monitor the likely accuracy of their source judgments and was the primary cause of their poor calibration scores.

As with the recognition analysis, we investigated whether it were possible that the foregoing differences were due to differences in the use of the confidence rating scale. However, there were no differences between either of the AD groups and the Older-m group in their overall distribution of confidence ratings when making source judgments, all t 's(22) < 1. These groups used the confidence scale in a similar manner and therefore, differences in the overall use of the confidence scale cannot explain the AD groups' propensity to make high confidence errors.

Overall, then, even though both groups of AD patients and the Older-m group are matched on overall source identification performance, they show different capacities to monitor the accuracy of these source judgments. Older-m adults show better monitoring resolution (i.e., better Somers' d correlations) and are more likely to assign higher confidence ratings to correct than incorrect source attributions. By contrast, both groups of AD patients are prone to make high confidence source identification errors.

3. General discussion

This study examined the ability of AD patients and healthy older adults to monitor and judge the likely accuracy of recognition judgments and source judgments about who spoke something earlier. Participants listened to statements at encoding that were presented by a woman and a man, and at a subsequent test phase, participants provided confidence ratings about the likely accuracy of their recognition judgments (i.e., was this statement encountered during the encoding phase or not?) and source judgments (i.e., was this statement presented by the man or the woman?). Consistent with prior findings, we observed that AD patients showed worse recognition performance and source identification performance than did

a group of older adults who experienced the identical study and test conditions (e.g., Budson et al., 2006; Dalla Barba et al., 1999; Multhaup & Balota, 1997). Moreover, repetition of the study material greatly improved recognition performance on the part of the AD patients but had little effect on source identification performance. This result is consistent with previous suggestions that recollection may be more disrupted than familiarity in AD patients to the extent that our old–new recognition test is influenced by a mixture of familiarity and source recollection whereas our source identification test is influenced primarily by source recollection since both sources were equally familiar (e.g., Budson et al., 2000; Dalla Barba, 1997; Gallo, Sullivan, Daffner, Schacter, & Budson, 2004; Knight, 1998; Souchay & Moulin, 2009; Westerberg et al., 2006). In addition, we observed that both groups of AD patients were very much impaired at monitoring the accuracy of both recognition and source identification judgments.

There are two striking findings about both groups of AD patients. First, they were worse at monitoring the likely accuracy of recognition and source judgments even when compared to a group of older adults (i.e., the Older-m group) who showed comparable recognition and source identification accuracy. For instance, for both recognition and source judgments Table 4 shows that the average confidence rating assigned by the AD-m patients to correct and incorrect judgments was nearly identical – indicating a nearly complete lack of awareness for the likely accuracy of correct and incorrect responses. By contrast, older adults on both tests were much more confident about the likely accuracy of correct than incorrect responses. The fact that these monitoring differences exist between the two groups – the AD-m group and the Older-m group – that show comparable recognition and source identification accuracy means that the AD-m patients' monitoring problems are not caused by a disproportionate difficulty remembering the information because there were no differences in accuracy between these older adults and AD patients. The second striking finding is that even when AD patients were given extra exposure to the study material, which produced a dramatic improvement in recognition performance (i.e., in Table 2 compare the recognition scores between the AD-m and the AD groups), there was little improvement in their ability to use confidence ratings to monitor the likely accuracy of these recognition judgments. For instance, Table 4 shows that the average confidence rating assigned by the AD-m and the AD groups to Hits and Misses was nearly identical. Finally, for both recognition judgments and source judgments, our results indicate that the monitoring impairment in AD patients is actually worse than their memory impairment (or at least conditions that are able to improve AD patients' memory are not able to improve monitoring performance), as otherwise there would have been no differences between the healthy older and the AD groups in monitoring performance when there were no differences in accuracy.

However, there is an interesting pattern on the part of the AD patients in that they show a severe monitoring impairment at the level of the item but not so much at the level of the task. As we have reviewed, at the level of the individual response, both groups of AD patients appear nearly completely unaware of the likely accuracy of a particular response. By contrast, at the level of the task, AD patients seem aware that the recognition judgment is an easier task than the source identification judgment. Specifically, both groups of AD patients were more accurate on the recognition task than on the source identification task and they provided significantly higher confidence ratings for the recognition task than for the source task. Two conclusions can be drawn from this pattern. First, AD patients do not show complete memory-monitoring anosognosia. Second, because confidence is tracking accuracy at the level of the task we can conclude that AD patients are not using the confidence rating scale in a haphazard manner and show some awareness that a higher confidence rating means a higher likelihood of being correct.

But, it remains for future research to investigate why AD patients show this memory-monitoring dissociation between responses to particular items and overall responses to one task or another.

Our study suggests that there are two different mechanisms that contribute to the monitoring impairment in AD patients. First, AD patients appear to have a deficit in monitoring episodic memory judgments that contributes to their worse monitoring performance (e.g., worse Somers' *d* scores) on both the recognition and the source identification tasks. Second, AD patients appear particularly vulnerable to making high-confidence errors on the source identification task that additionally contributes to their monitoring impairment on this task.

To explain the monitoring behavior of Alzheimer's patients we propose a Remembrance–Evaluation model of monitoring. This is a two-stage model that builds on the ideas from Johnson et al. (1993)' Source-Monitoring Framework, Schacter, Norman, and Koutstaal (1998)' Constructive Memory Framework, and Koriat (1993)'s Accessibility Model of monitoring. The first, Remembrance, stage of the model refers to remembered information, and as in the models of Johnson, Schacter, Koriat and their respective colleagues, a confidence rating or a monitoring judgment is based, in part, on the kind and amount of remembered information. For our purposes, remembered information can lead to monitoring failures primarily when the remembered information is illusory and false but it elicits the same kind of subjective experience as true memories that are assigned high confidence responses. These high confidence misrecollections will affect monitoring performance on recollective-based tests, such as cued-recall or source memory tests. We suggest that it is these misrecollections that are causing the occurrence of high-confidence errors by the AD patients on the source judgment.

The second stage of our model is the evaluation process. Once information is remembered, it is necessary to consciously evaluate this information with the purpose of both making a memory judgment (e.g., does the kind and amount of remembered information allow for a judgment of 'old') and making a confidence rating (or an analogous assessment) about the likely accuracy of this memory judgment. Consequently, this evaluation stage can produce monitoring failures on different kinds of memory tests when individuals use inappropriate criteria to evaluate the remembered information, such as inappropriately weighting the presence or absence of particular kinds of memorial information. This inappropriate evaluation will in turn produce an inappropriate judgment (e.g., confidence rating) about the likely accuracy of a response. To simplify the model, we are assuming that a malfunctioning evaluative process will comparably affect the evaluation of different kinds of memorial information (e.g., familiarity and perceptual information). However, given the studies showing that AD patients appear normal at monitoring the accuracy of responses to general knowledge questions (e.g., Backman & Lipinska, 1993), a distinction should be made between episodic and semantic information with AD patients showing preserved evaluative processes of semantic information and impaired evaluative processes of episodic information. Additional support for this episodic–semantic distinction comes from Reggev, Zuckerman, and Maril (2011) who show that different brain regions are associated with monitoring these different kinds of memory. The key point about a malfunctioning evaluation process is that it can contribute to a pattern of impaired monitoring on different kinds of episodic memory tests, which would produce the monitoring impairment by the AD patients on both the recognition and source memory judgments.

Overall, then, the crux of the Remembrance–Evaluation model is that monitoring impairments can be caused by two very different mechanisms: (1) remembered information that is distorted and illusory; and (2) evaluative processes that are inappropriate. The particular signature of the monitoring impairment will depend on which or both of these mechanisms that is malfunctioning.

Consider the application of Remembrance–Evaluation model to the monitoring behavior of healthy older adults and AD patients. Growing evidence suggests that, when compared to young adults, healthy older adults show a selective monitoring deficit on source memory tests and cued-recall tests that is caused by high confidence errors; but they show no monitoring deficit when evaluating recognition judgments (e.g., Dodson, Bawa, & Krueger, 2007). The selectivity of the monitoring deficit in older adults suggests that it is caused by a malfunctioning Remembrance mechanism and not by problems with the Evaluation mechanism. In other words, older adults evaluate remembered information normally; their monitoring deficit is caused by a tendency to remember false information. By contrast, AD patients' monitoring deficit shows a different signature: it is more widespread and occurs on both recognition and source identification judgments. Moreover, because (1) we observed this monitoring deficit on both recognition judgments and source judgments and, (2) given the general assumption that recognition judgments are more familiarity-based than are source judgments, we see no evidence from our study for the notion that AD patients show relatively preserved monitoring of familiarity information and impaired monitoring of recollection information (Souchay, 2007). Overall, then, we suggest that both stages of the Remembrance–Evaluation model are malfunctioning in AD patients. That is, AD patients are prone to falsely recollect information that leads to high confidence errors and they are impaired at evaluating remembered information in order to provide a confidence rating about its likely accuracy.

Consideration of the brain pathology in AD, and more generally the neural correlates of memory-monitoring, may aid our understanding of why both stages of the Remembrance–Evaluation model are faulty in AD patients. In addition to the medial temporal lobes, AD patients also show pathology in parietal (McKee et al., 2006) and frontal (Lidstrom et al., 1998) cortex. Functional neuroimaging studies have documented a relationship between all of these foregoing brain areas and monitoring the accuracy of memory (e.g., Chua, Schacter, Rand-Giovannetti, & Sperling, 2006). However, patient studies may provide more direction about the role of particular brain regions in this monitoring process. There is increasing evidence that the parietal lobes may underlie the subjective memorial experience of recollection (e.g., Ally, Simons, McKeever, Peers, & Budson, 2008; Davidson et al., 2008; Simons et al., 2010). Simons et al. (2010) collected confidence ratings by patients with unilateral or bilateral lesions of the parietal lobe and controls with a combined recognition and source recollection paradigm that is nearly identical to our task. They observed no differences between either group of parietal patients and controls on either recognition accuracy or on the overall average confidence rating associated with the recognition judgment. By contrast, even though there were no differences between the parietal patients and controls on source recollection performance, the patients with bilateral parietal lesions showed significantly lower confidence ratings than controls in their source identification judgments. These results are consistent with their subjective recollection hypothesis that the parietal lobes contribute to the recovery and assessment of details that contribute to recollective experience – and, which in turn contribute to confidence ratings for recollective judgments. However, Simons et al.'s observations of diminished confidence on the source identification task that accompanies bilateral parietal damage is the opposite of what we have observed in our AD patients: in contrast to diminished confidence, our AD patients show excessive high confidence errors on the nearly identical source recollection task. With respect to our Remembrance–Evaluation model, the Simons et al. study suggests that the bilateral parietal patients show preserved evaluation processes – which account for their preserved monitoring performance on the recognition task – but an impaired

and diminished remembrance process that accounts for their reduced confidence ratings on the source task.

While, to our knowledge, there are no studies involving patients with frontal lobe damage that have used a combined recognition/source-identification and confidence task, there is growing evidence from other tasks that suggests the involvement of the frontal lobes in metamemorial judgments. For instance, frontal patients are worse than matched controls at providing accurate feeling of knowing (FOK) judgments about the likelihood of recognizing a not-recalled answer on an episodic cued-recall task (e.g., Janowsky, Shimamura, & Squire, 1989; Schnyer et al., 2004). Interestingly, the worse monitoring performance by the frontal lobe patients tends to occur because of overconfidence. For instance, Janowsky et al. observed that frontal patients were less likely than controls to correctly recognize unrecalled items when both groups were either moderately or highly confident that they would be able to correctly recognize the item. This pattern of overconfidence is similar to the overconfidence that we have observed in our AD patients. By contrast, patients with medial temporal lobe damage are no different from matched controls in the accuracy of their FOK judgments (see Pannu & Kaszniak, 2005 for review).

Overall, then, there are multiple ways that the frontal lobes could contribute to malfunctioning Remembrance and Evaluation processes. According to Schacter and colleagues' constructive memory framework, the frontal lobes contribute both to retrieval-cue specification (i.e., focusing) processes and to evaluation processes (Schacter et al., 1998). Poor cue specification can activate memories that are either not appropriate for the target task or are inappropriately vague. In addition, inappropriate evaluation criteria when judging memories have been used to explain the pathologically high false recognition rates by frontal patients (e.g., Schacter, Curran, Galluccio, Milberg, & Bates, 1996). Both of these processes can explain our findings. AD patients may not be specifying the appropriate retrieval cues, which may cause them to misremember past events – thus leading to high-confidence errors on the source task. Moreover, a malfunctioning evaluation process may cause the AD patients to inappropriately use the confidence rating scale on both recognition and source judgment tasks. The AD patients' frontal lobe pathology likely explains much of our findings.

The malfunctioning of the Evaluation stage of the Remembrance–Evaluation model may also help to explain the abnormally liberal recognition response bias seen in patients with AD. It has been shown previously that AD patients exhibit a more liberal response bias compared with healthy older adults (Budson et al., 2006), although the patients are able to shift to a more conservative bias when given an instructional manipulation (such as being told that only 30% of test items have been studied) (Waring, Chong, Wolk, & Budson, 2008). Importantly for the present study, Budson et al. (2006) found that the recognition response bias of AD patients remained more liberal than that of healthy older adults even when their recognition memory performance was equated by varying the length of study and test lists. Although the factors that contribute to recognition response bias presumably involve both conscious evaluation and unconscious processes, that patients with AD showed impaired evaluation of their memory in the present experiment may provide an important clue as to why patients with AD show an abnormally liberal response bias. We found that both groups of AD patients were comparably confident about the accuracy of both their correct and incorrect recognition responses to studied items, in contrast to the control groups which each showed higher rates of confidence for their correct than for their incorrect responses. Future studies can work towards examining both memorial confidence and recognition response bias to better understand the relationship between these processes in patients with AD.

It is important to note that the monitoring results from our recognition judgment (i.e., worse monitoring performance by the AD patients than older controls) conflict with the findings of Moulin et al. (2003) who observed that AD patients were no different from healthy older adults in monitoring the accuracy of recognition judgments. However, a fundamental difference between our recognition task and that of Moulin et al. is that they used a 2 alternative, forced-choice procedure in which participants were presented simultaneously with an old event and a new event. It likely is easier to assess the accuracy of a recognition judgment when one is evaluating a previously encountered event in relation to a novel event that is simultaneously present on the screen. That is, one's confidence that a recognition judgment is correct can be determined by both the certainty that one event is old as well as the certainty that the other event is novel – a quasi-triangulation method for determining the likely accuracy of a recognition judgment. By contrast, fewer cues are available for making this monitoring judgment in our recognition procedure in which a single event is presented and participants endorse the event as old or new and then assess the likely accuracy of this judgment. Thus, this difference in recognition results between our study and Moulin et al. suggests that the evaluation stage on the part of AD patients can be facilitated by providing additional cues or support at retrieval for making a memory judgment, as in Moulin et al.'s study.

There are a couple of potential limitations to our study and points to consider for future research. First, there are a small number of items that each individual contributes to the overall data (i.e., only 30 items presented at encoding to each participant). Although the small number of items is a limitation, it should be noted that our central effects were replicated within each AD group and within each older adult group, which testifies to the reliability of these effects. Second, we used different matching procedures in our AD-m group (i.e., extra repetitions of the study material) and in our Older-m adult group (i.e., adding a large number of filler items at encoding to increase study list length) in order to match accuracy and thus remove differences in accuracy as a confounding explanation for differences in monitoring performance. However, is it possible that these matching manipulations have introduced a confound by changing the cognitive processes associated with the monitoring judgment? In answer to this question it is critical to point out that the older group that received a matching manipulation showed a pattern of monitoring performance that was nearly identical to the pattern shown by the older adult group that did not receive a matching manipulation. The same observation holds for the AD-m and the AD group. Because the same pattern of monitoring performance occurs within each group of older adults and within each group of AD patients, it is reasonable to conclude that the difference in monitoring ability between both groups of AD patients and both groups of older adults is caused by some variable related to Alzheimer's disease and is not caused either (1) by differences in memory strength since accuracy was equated or (2) by the matching manipulations.

A point to consider in future research is the effect of our instructions to participants to try to use the entire scale of confidence ratings over the course of the experiment. This instruction is common in confidence-rating studies because it is intended to orient participants to use and interpret the confidence rating scale in a similar manner and thus to minimize biases to either use one end of the scale or the middle points only or the end-points only, etc. Our analyses of the distribution of confidence ratings for each judgment – regardless of accuracy – suggest that there were no differences in the overall use of the scale between our AD patients and the healthy older adults. However, it is a question for future research whether AD patients would show better confidence–accuracy correspondence when they are not given these instructions to 'use the entire scale.'

Finally, our neuropsychological tests showed that the individuals in the AD-m group appeared less impaired than those in the AD group, as the AD-m individuals showed significantly better MMSE and TMT-B scores and were numerically better on all of the other neuropsychological tests. We suggest that there are important theoretical implications to this group difference to the extent that it is the case that this reduced cognitive impairment is due to the AD-m individuals, on average, showing an earlier stage of the disease. Despite experiencing an earlier stage of the disease, the AD-m individuals show a devastating monitoring impairment that in many ways is just as severe as the monitoring impairment by the AD group. Given (1) that overall objective memory performance is matched between the AD-m and the Older-m group and (2) the apparent early stage of the disease, we therefore suggest that the episodic memory monitoring impairment that we have documented may manifest itself in AD patients before a corresponding impairment in memory performance (see Galeone, Pappalardo, Chieffi, Iavarone, & Carlomagno, 2011 for a related argument).

In conclusion, we observed a dissociation between AD patients' recognition and source accuracy on the one hand, and their ability to monitor and assess the likely accuracy of their recognition and source judgments on the other. Strikingly, the AD patients' monitoring impairment persists even when their memory accuracy is boosted to that of a healthy older group which indicates that the AD's monitoring difficulties are not caused by difficulties remembering the information. There are serious practical consequences to this kind of monitoring impairment. Our finding that AD patients are prone to make high-confidence source errors indicates that they may be particularly vulnerable to disrupted medication regimens in which they either fail to take or repeatedly take medication because they are highly confident in the accuracy of the memory that supports this action. Similarly, patients may be highly confident in their memory that they have already turned the stove off when they have not. In addition, we speculate that the high-confidence source errors that we have observed on the part of our AD patients are similar to the confabulatory responses that are exhibited by these individuals and that consequently there may be a common mechanism underlying both behaviors. We are hopeful that studies such as ours that evaluate memorial confidence can lead to future behavioral and pharmacological interventions that may lead to patients' improved ability to question and probe their own memories prior to acting. Such interventions may be able to improve the lives of patients and their families.

References

- Allen-Burge, R., & Storandt, M. (2000). Age equivalence in feeling-of-knowing experiences. *Journal of Gerontology: Psychological Sciences*, *55B*, P214–P223.
- Ally, B. A., Simons, J. S., McKeever, J. D., Peers, P. V., & Budson, A. E. (2008). Parietal contributions to recollection: Electrophysiological evidence from aging and patients with parietal lesions. *Neuropsychologia*, *46*, 1800–1812.
- Backman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease. *Neuropsychologia*, *31*, 335–345.
- Brewer, N., & Wells, G. L. (2006). The confidence–accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11–30.
- Budson, A. E., Daffner, K. R., Desikan, R., & Schacter, D. L. (2000). When false recognition is unopposed by true recognition: Gist based memory distortion in Alzheimer's disease. *Neuropsychology*, *14*, 277–287.
- Budson, A. E., Dodson, C. S., Daffner, K. R., & Schacter, D. L. (2005). Metacognition and false recognition in Alzheimer's disease: Further exploration of the distinctiveness heuristic. *Neuropsychology*, *19*, 253–258.
- Budson, A. E., Simons, J. S., Sullivan, A. L., Beier, J. S., Solomon, P. R., Scinto, L. F., et al. (2004). Memory and emotions for the September 11, 2001, terrorist attacks in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Neuropsychology*, *18*, 315–327.
- Budson, A. E., Simons, J. S., Waring, J. D., Sullivan, A. L., Hussion, T., & Schacter, D. L. (2007). Memory for the September 11, 2001, terrorist attacks one year later in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Cortex*, *43*, 875–888.
- Budson, A. E., Wolk, D. A., Chong, H., & Waring, J. D. (2006). Episodic memory in Alzheimer's disease: Separating response bias from discrimination. *Neuropsychologia*, *44*, 2222–2232.
- Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*, 654–663.
- Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: Neural correlates of the cognitive process and subjective level of confidence in recognition memory. *NeuroImage*, *29*, 1150–1160.
- Cosentino, S., Metcalfe, J., Butterfield, B., & Stern, Y. (2007). Objective metamemory testing captures awareness of deficit in Alzheimer's disease. *Cortex*, *43*, 1004–1019.
- Dalla Barba, G. (1997). Recognition memory and recollective experience in Alzheimer's disease. *Memory*, *5*, 657–672.
- Dalla Barba, G., Nedjam, Z., & Dubois, B. (1999). Confabulation, executive functions and source memory in Alzheimer's disease. *Cognitive Neuropsychology*, *16*, 385–398.
- Davidson, P. S. R., Anaki, D., Ciaramelli, E., Cohn, M., Kim, A. S. N., Murphy, K. J., et al. (2008). Does lateral parietal cortex support episodic memory? Evidence from focal lesion patients. *Neuropsychologia*, *46*, 1743–1755.
- Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory and high confidence errors: A misrecollection account. *Psychology and Aging*, *22*, 122–133.
- Dodson, C. S., Bawa, S., & Slotnick, S. D. (2007). Aging and a signal detection model of illusory recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 169–181.
- Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: Why older adults are unreliable eyewitnesses. *Psychological Bulletin & Review*, *13*, 770–775.
- Galeone, F., Pappalardo, S., Chieffi, S., Iavarone, A., & Carlomagno, S. (2011). Anosognosia for memory deficit in amnesic mild cognitive impairment and Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *26*, 695–701.
- Gallo, D. A., Sullivan, A. L., Daffner, K. R., Schacter, D. L., & Budson, A. E. (2004). Associative recognition in Alzheimer's disease: Evidence for impaired recall-to-reject. *Neuropsychology*, *18*, 556–563.
- Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, *38*, 771–784.
- Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology*, *17*, 3–11.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.
- Knight, R. G. (1998). Controlled and automatic memory processes in Alzheimer's disease. *Cortex*, *34*, 427–435.
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*, 609–639.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517.
- Lidstrom, A. M., Bogdanovic, N., Hesse, C., Volkman, I., Davidsson, P., & Blennow, K. (1998). Clusterin (apolipoprotein J) protein levels are increased in hippocampus and in frontal cortex in Alzheimer's disease. *Experimental Neurology*, *154*, 511–521.
- Lipinska, B., & Backman, L. (1996). Feeling-of-knowing in fact retrieval: Further evidence for preservation in early Alzheimer's disease. *Journal of the International Neuropsychological Society*, *2*, 350–358.
- McKee, A. C., Au, R., Cabral, H. J., Kowall, N. W., Seshadri, S., Kubilus, C. A., et al. (2006). Visual association pathology in preclinical Alzheimer disease. *Journal of Neuropathology and Experimental Neurology*, *65*, 621–630.
- Mitchell, J. P., Sullivan, A. L., Schacter, D. L., & Budson, A. E. (2006). Misattribution errors in Alzheimer's disease: The illusory truth effect. *Neuropsychology*, *20*, 185–192.
- Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., Van Belle, G., Fillenbaum, G., et al. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD). *Neurology*, *39*, 1159–1165.
- Moulin, C. J. A., James, N., Perfect, T. J., & Jones, R. W. (2003). Knowing what you cannot recognize: Further evidence for intact Metacognition in Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, *10*, 74–82.
- Multhaup, K. S., & Balota, D. A. (1997). Generation effects and source memory in healthy older adults and in adults with dementia of the Alzheimer type. *Neuropsychology*, *11*, 382–391.
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, *15*, 105–130.
- Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology*, *14*, 330–345.
- Pierce, B. H., Sullivan, A. L., Schacter, D. L., & Budson, A. E. (2005). Comparing source-based and gist-based false recognition in aging and Alzheimer's disease. *Neuropsychology*, *19*, 411–419.
- Pierce, B. H., Waring, J. D., Schacter, D. L., & Budson, A. E. (2008). Further examination of recall-to-reject processes in aging and Alzheimer's disease. *Cognitive and Behavioral Neurology*, *21*, 179–186.
- Pliske, R. M., & Mutter, S. A. (1996). Age differences in the accuracy of confidence judgments. *Experimental Aging Research*, *22*, 199–216.
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, C. D., et al. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, *15*, 1676–1689.

- Reggev, N., Zuckerman, M., & Maril, A. (2011). Are all judgments created equal? An fMRI study of semantic and episodic metamemory predictions. *Neuropsychologia*, 49, 1332–1342.
- Schacter, D. L., Curran, T., Galluccio, L., Milberg, W. P., & Bates, J. F. (1996). False recognition and the right frontal lobe: A case study. *Neuropsychologia*, 34, 793–808.
- Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neurosciences of constructive memory. *Annual Review of Psychology*, 49, 289–318.
- Schnyer, D. M., Verfaellie, M., Alexander, M., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, 42, 957–966.
- Shimamura, A. P., & Wickens, T. D. (2009). Superadditive memory strength for item and source recognition: The role of hierarchical relational binding in the medial temporal lobe. *Psychological Review*, 116, 1–19.
- Simons, J. S., Peers, P. V., Mazuz, Y. S., Berryhill, M. E., & Olson, I. R. (2010). Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. *Cerebral Cortex*, 20, 479–485.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34–50.
- Souchay, C. (2007). Metamemory in Alzheimer's disease. *Cortex*, 43, 987–1003.
- Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, 14, 299–309.
- Souchay, C., Isingrini, M., & Gil, R. (2002). Alzheimer's disease and feeling-of-knowing in episodic memory. *Neuropsychologia*, 40, 2386–2396.
- Souchay, C., & Moulin, C. J. A. (2009). Memory and consciousness in Alzheimer's disease. *Current Alzheimer Research*, 6, 186–195.
- Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279–306.
- Tendolkar, I., Schoenfeld, A., Golz, G., Fernandez, G., Kuhl, K. P., Ferszt, R., et al. (1999). Neural correlates of recognition memory with and without recollection in patients with Alzheimer's disease and healthy controls. *Neuroscience Letters*, 263, 45–48.
- Waring, J. D., Chong, H., Wolk, D. A., & Budson, A. E. (2008). Preserved metamemorial ability in patients with Alzheimer's disease: Shifting response bias. *Brain & Cognition*, 66, 32–39.
- Westerberg, C. E., Paller, K. A., Weintraub, S., Mesulam, M. M., Holdstock, J. S., Mayes, A. R., et al. (2006). When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer's diseases. *Neuropsychology*, 20, 193–205.
- Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110, 611–617.