Published in final edited form as:

Neuropsychology. 2012 July ; 26(4): 509–516. doi:10.1037/a0028684.

Memory Distortion in Alzheimer's Disease: Deficient Monitoring of Short and Long-term Memory

Katherine E. MacDuffie^{1,*}, Alexandra S. Atkins¹, Kristin E. Flegal³, Christopher M. Clark⁴, and Patricia A. Reuter-Lorenz²

¹Duke University

²University of Michigan

³University of California, Davis

⁴University of Pennsylvania

Abstract

Objective—This study measured distortions of memory during short-term memory (STM) and long-term memory (LTM) versions of a semantically-associated word list learning paradigm. Performance of patients with mild-to-moderate Alzheimer's disease (AD; MMSE = 16) was compared to performance of age-matched, healthy older adult participants.

Method—In a STM version of the Deese-Roediger-McDermott (DRM) task, participants viewed four-word lists and were prompted for recall after a brief interval. The LTM task tested recall memory for 12-word lists.

Results—Compared to the healthy group, the AD participants show greater impairment on the LTM task than on the STM task, although veridical recall is significantly reduced on both tasks. Furthermore, on both memory tasks: a) Participants with AD generate more nonsemantic intrusions than healthy older adult participants; b) semantic intrusion rate, when computed as a proportion of total recall, does not differ between groups. Notably, nonsemantic intrusions are consistently high for AD participants across both STM and LTM despite a marked difference in recall accuracy (65% and 23%, respectively).

Conclusions—STM impairment with some preserved semantic processing is evident in AD. The extent and variety of intrusions reported by AD participants indicates a breakdown in their ability to monitor and constrain their recall responses, even within seconds of initial learning.

Keywords

Alzheimer's disease; false memory; short-term memory; long-term memory; semantic memory

Distortion of memory in Alzheimer's disease (AD) is a clinically relevant concern. Patients and their caregivers frequently report incidents such as getting momentarily lost on familiar routes, placing possessions in inappropriate locations, and confusing current context with past experiences—errors that reflect distortions of existing memories, rather than complete forgetting. In a study examining memory for the events of September 11th, 2001, patients with AD reported fewer correct memories in the days that followed the attacks as compared to control participants, and also showed more distortion of those memories in a follow-up interview three months later (Budson et al., 2004). Examining such distortion with

*Corresponding author: Duke University, Psychology and Neuroscience, 417 Chapel Drive, Box 90086, Durham, NC 27708, km170@duke.edu, 919-681-4601.

laboratory-controlled memories could provide clinically useful information about the nature, severity, and mechanisms of memory errors in AD. The present study takes a new approach to these issues by examining memory distortions under short-term memory conditions where the memory load and retention interval are minimal, thereby maximizing the potential for veridical, or accurate, memory of studied items.

The most widely used paradigm for studying memory distortion in the laboratory is the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). In this task, participants view a list of words, all of which bring to mind a strong semantic associate—the theme word—which is never presented. For example, when participants see the words: *dream, bed, nap, doze, awake, snooze*, the common theme word is *sleep*. In the original study (Roediger & McDermott, 1995), healthy college-aged participants recalled the unstudied theme word on 40% of trials, and falsely recognized the theme word as often as words that appeared in the middle of the list. These results demonstrate the vulnerability of memory to distortions due to associative processes. Indeed, false memories are generally thought to be due to activational processes that bring semantic associations to mind during encoding or retrieval, and monitoring processes at retrieval that serve to adjudicate between studied and unstudied items (Roediger et al., 2001).

A number of laboratories have investigated the performance of AD patients on the DRM and related false-memory paradigms, with somewhat inconsistent results. While some studies report a decrease in false recognition and recall with AD relative to healthy older adults (Budson et al., 2001; Budson et al., 2003), others have observed an equivalent or increased level of associative errors (Budson et al., 2000; Budson, Sitarski, et al., 2002; Budson, Sullivan, et al., 2002; Gallo et al., 2004; Pierce et al., 2005), especially when examined in relation to veridical memory (Balota et al., 1999; Watson et al., 2001). These results have been differentially attributed to deficits in item-specific recollection in AD (Budson et al., 2000; Budson et al., 2003; Gallo et al., 2006), degraded gist representations (e.g., Brainerd & Reyna, 2005; Budson et al., 2001), or impaired monitoring (e.g., Johnson et al., 1993; Budson, Sitarski, et al., 2002; Budson, Sullivan, et al., 2002; Gallo et al., 2004; Pierce et al., 2005).

Item-specific recollection refers to the retrieval of memory representations that include the specific details, such as the perceptual format or context, of the encoding episode. These details can be used to confirm that the item in question was indeed part of the stimulus set (Johnson et al., 1993; McCabe et al., 2004). In a false memory paradigm, item-specific recollections are only available for studied items, not for their unstudied semantic associates, and thus can be used to distinguish studied from unstudied items at retrieval (e.g., Schacter et al., 1999). Gist, on the other hand, is the representation of the general semantic meaning of a stimulus, and is not tied to the specific experience of viewing the memory set (e.g., Brainerd & Reyna, 2005). Thus, gist exists for both studied and unstudied semantic associates in a DRM paradigm. According to one prominent account of false memories, known as Fuzzy Trace Theory (Brainerd and Reyna, 2005) gist and item-specific (also referred to as *verbatim*) traces can be stored in parallel, but come into competition at the time of a retrieval decision. Accurate performance on a DRM task presumably relies on monitoring processes that enable participants to distinguish the items supported by gist *and* verbatim traces, in order to endorse only studied items at test. The two traces also display distinct decay patterns—while gist traces tend to be durable over time, verbatim traces degrade more quickly, and so would be expected to be more available at short than long delays. The differing temporal profiles of the two traces may mean that manipulating the delay between encoding and retrieval in a DRM task could alter their relative strength at the time of a retrieval decision (but see Flegal et al., 2010).

While there is abundant evidence that monitoring is impaired in AD (Dalla Barba et al., 1999; Budson, Sullivan, et al., 2002; Budson, Sitarski, et al., 2002; Gallo et al., 2004; Gallo et al., 2007), the contribution of such deficits to poor memory performance is uncertain. Monitoring deficits are generally linked to frontal lobe dysfunction (Butler et al., 2004), and monitoring difficulties in AD may be attributable to disease-related frontal lobe pathology (Budson, Sullivan, et al., 2002; Mitchell et al., 2006). The clearest behavioral manifestation of impaired monitoring in AD is an increased intrusion rate on recall tasks, which has been reported in number of studies (Balota et al., 1999; Watson et al., 2001), and has even been examined as a possible staging tool or diagnostic marker for the disease (Lowenstein et al., 1991; Rouleau et al., 2001; Fuld et al., 1982).

Two previous studies highlight the inability of Alzheimer's patients to effectively monitor and constrain their responses on variants of the DRM paradigm. Budson and colleagues (2000) examined the influence of repeated encoding of DRM lists with young adult, healthy older adult, and AD participants. Results showed that increasing the number of encoding instances increased true recognition for all subject groups. For young and healthy older adult controls, this increase in veridical memory was accompanied by a decrease in false memory. For AD patients, however, both true and false memory increased with encoding repetitions. In a subsequent study with AD patients, Pierce et al. (2005) used a deep vs. shallow encoding manipulation in an effort to increase item-specific recollection for deeply encoded items. As the authors expected, deep encoding lead to increased veridical memory and reduced false memory for healthy controls. For AD participants, however, false memory rates increased following deep encoding.

The authors of these studies suggest that while healthy participants were able to use the strengthened item-specific representations for the studied words to more accurately monitor their responses at retrieval, AD participants were not. Instead, as the semantic gist accumulated over encoding repetitions and with deep encoding, AD participants appeared unable to discriminate between the strengthened gist trace and the items in the study set. Thus, the authors conclude that dysfunctional memory monitoring is responsible for the elevated levels of false recognition by AD participants in these paradigms (Budson et al., 2000; Budson, Sullivan et al, 2002; Pierce et al., 2005).

Our study aims to test this interpretation of memory distortions in AD using a task that enhances item-specific memory in a different way: by radically reducing the memory load and the retention interval. To test the vulnerability of short-term memory to semantic distortions, Atkins and Reuter-Lorenz (2008) developed a short-term memory version of the DRM paradigm. They demonstrated that healthy young adults generate false semantic memories when asked to recall lists of only four associated items following a 4-second retention interval (Atkins and Reuter-Lorenz, 2008). The current study used this short-term memory DRM task to test individuals diagnosed with AD, and healthy age-matched older adults. We compare performance of the groups on both a short-term memory (STM; 4-word) and on a long-term memory (LTM; 12-word) version of the DRM task.

The reduced memory demands (fewer items to remember for briefer intervals) in the STM DRM task should provide optimal conditions for individuals to overcome false memories relative to the traditional DRM task. Item-specific information and relatively rich contextual detail should be readily available to support veridical memory and to counteract the distorting influence of gist, increasing the likelihood that individuals with AD will utilize these cues if they retain the ability to do so. This might lead us to expect that differences between individuals with and without AD will be minimal in the STM DRM task relative to the LTM task. Note however that the effects of these task modifications on gist processing are uncertain. On the one hand, the reduced number of semantically related items, coupled

with less time for elaboration, could also result in reduced gist and a lower false memory rate for the AD group relative to healthy older adults. That is, the opportunity for semantic association may be too limited for the weakened gist-processing system in AD (Budson et al., 2001; Budson et al., 2003; Gallo et al., 2006), resulting in very few or no false memory intrusions for AD participants in the STM task. On the other hand, better memory for list items in the STM task could result in strengthened gist (as in the aforementioned studies; Budson et al., 2000; Pierce et al., 2005), which could lead to more false memories for AD participants compared to healthy older adults, and compared to their own performance in the LTM version of the task. In either case, the present design provides a novel approach for assessing STM function (e.g., Belleville et al., 1996; Kensinger, et al., 2003; Peters and Collette et al., 2009; Peters and Majerus et al., 2009) and the causes of memory distortions in AD.

Material and methods

Participants

19 participants with AD (9 female, mean age = 77.5 years ($SD = 5.95$)) were recruited from the Penn Memory Center at the University of Pennsylvania. All patients had a diagnosis of Probable AD, determined by the National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorder Association criteria (NINCDS-ADRDA; McKhann et al., 1984), a Mini Mental State Evaluation (MMSE; Folstein, Folstein & McHugh, 1975) score of 16 or higher (mean score 21.61 ($SD = 3.66$), range: 16–30), and no evidence of depression or other neurological impairments. All study procedures were approved with regards to human subjects protection by the University of Pennsylvania Institutional Review Board, and written informed consent was obtained from all AD participants and their caregivers. 19 older adult (OA) control participants (10 female, mean age = 74 years ($SD = 5.63$)) were recruited from the University of Michigan. OA participants were screened for dementia and the other exclusion criteria mentioned above (mean MMSE score = 28.68 ($SD = 1.42$), range: 25–30). All study procedures were approved with regards to human subjects protection by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board, and written informed consent was obtained from all OA participants. The ages of the two groups did not differ significantly ($F(1, 36) = 2.821, p > .1, \eta^2 = .07$), and MMSE scores of the AD group were significantly lower than those of the OA group ($F(1, 36) = 61.22, p < .001, \eta^2 = .64$). Note that the single individual diagnosed with probable AD who had an MMSE of 30, scored below the 10th percentile in logical memory and had biomarkers consistent with AD. The results of all data analyses are unchanged if this individual is excluded.

Materials and Procedure

The stimuli for the false memory tasks consisted of 26 lists of words. Each list consisted of associated words that were related in meaning to a nonpresented theme word. Each list was associated with a unique theme word. Six lists (containing 12 words each) were taken from Roediger and McDermott (1995), and were used for the LTM recall task. The remaining 20 lists contained four words each and were used for the STM task. These were selected from a set of 128 four-word lists developed for false short-term memory research as previously described (Flegal et al., 2010). An example STM list for the theme word SOAP is: LATHER, SUDS, SHOWER, BUBBLE. These lists match the range of values of the original DRM lists for frequency and concreteness, and have been shown to reliably elicit false memories in recognition and recall tests of both STM and LTM (Flegal et al., 2010). The lists had no items in common and were never repeated. The LTM lists had higher average Backward Associative Strength (BAS) than the STM lists (.486 and .405, respectively). BAS is a measure of the association strength of a list item back to the critical

item (i.e. likelihood of recalling critical item SLEEP when presented with list item BED). BAS values for the LTM lists were copied from Roediger et al. (2001), and values for the STM lists were taken from the University of South Florida Free Association Norms (Nelson, McEvoy & Shreiber, 1998). The fact that our LTM lists had higher average BAS than the STM lists could result in higher rates of semantic intrusions for the LTM lists (Robinson & Roediger, 1997). Because of this difference, and other task differences noted below, our quantitative analyses of errors do not involve direct comparisons of STM and LTM intrusions. Instead, we compare the quantity of each intrusion type between AD and OA groups for each task separately.

The tasks were administered on a computer using E-prime 1.1 Software (Psychology Software Tools, Pittsburgh, PA). Participants' recall responses were recorded with a handheld digital voice recorder and transcribed after the session. Responses were coded as either correct (i.e., present in the memory set) or incorrect (regardless of output order), using previously established criteria (Atkins & Reuter-Lorenz, 2008). Specifically, incorrect responses were classified into one of three categories. 1) *Semantic* intrusions included both nonpresented theme words and words that related in meaning to at least two of the words in the memory set. 2) *Phonological* intrusions were responses that rhymed with one of the presented words. Finally, 3) *Other* intrusions were perseverative responses from previous memory sets, non-word utterances, and unrelated intrusions. Note that within-list perseverations were not included in these tallies because participants were not explicitly warned against making these responses, and experimenter observation suggested that such responses were used primarily by OAs to self-cue in an effort to retrieve more items from memory. The overall higher mean frequency of within-list perseverations for the OA than for the AD group (approximately .65 versus .02 respectively) supports our interpretation of these responses and our decision not to score within-list perseverations as errors.

Short-Term Memory Recall

In the STM recall task, a list of four semantically related words appeared on the screen simultaneously and were read aloud by participants at their own pace, which typically ranged between 4–10 seconds. The words were followed by a 3–4 second retention interval, during which participants counted aloud. A recall screen then prompted participants to say the words they remembered from the beginning of the trial. Because this task emphasizes short-term memory storage, and does not include the manipulation of items or additional processing of stored information, as is typical of working memory tasks like reading or operation span, we refer to this as a measure of STM, and not working memory per se (see e.g., Baddeley, 1986; Jonides et al., 2008; Reuter-Lorenz & Jonides, 2007). However, the use of articulatory suppression presumably interferes with exclusive reliance on the phonological loop and may promote reliance on visual or semantic codes.

Long-Term Memory Recall

In the LTM recall task, participants viewed 6 lists of 12 words each. The words in each list were presented individually on a computer screen for 2 seconds, with a 1 second interval between words. Participants read the words out loud as they appeared, and each list was immediately followed by a recall prompt.

Thus, the LTM task differed from the STM task in list length and procedurally in that the STM lists were presented simultaneously, and followed by articulatory suppression. We opted to settle for these task differences in order to follow procedures for each task that had been shown previously to be effective in eliciting false memories. This was especially important for the STM task because false recall errors under short-term task conditions have been relatively unexplored and in our hands to date had only been demonstrated with

simultaneous presentation (Atkins & Reuter-Lorenz, 2008; Coane et al. 2007). We therefore made only minor modifications in this procedure to accommodate the needs of the AD group.

Results

To facilitate comparison of STM and LTM accuracy, percent correct recall was calculated for each subject in each task by dividing the mean number of correct responses by the total number of items in the memory set (STM = 4, LTM = 12). Mean ($\pm$ SD) correct recall for STM and LTM respectively, are as follows: 65.2% ($\pm$ 17.0%) and 23.0% ($\pm$ 11.5%) for the AD group; 84.2% ($\pm$ 7.0%) and 53.5% ($\pm$ 8.1%) for the OA group. A 2×2 group $\times$ task analysis of variance (ANOVA) on percent correct recall revealed a main effect of group (OA > AD; $F(1, 36) = 51.09$, $p < .001$, $\eta^2 = .59$), a main effect of task (STM > LTM; $F(1, 36) = 624.80$, $p < .001$, $\eta^2 = .95$), and a significant group $\times$ task interaction ($F(1, 36) = 15.63$, $p < .001$, $\eta^2 = .3$). These results indicate that OA participants have better veridical memory than the AD group in both the LTM and STM tasks. While percent correct recall is lower on the LTM task for both groups, the interaction effect suggests that AD participants are disproportionately impaired on this task in comparison to healthy OAs. The effect size associated with the group difference in LTM ($\eta^2(36) = 9.47$, $p < .001$, $d = 3.07$) is more than double the effect size in STM ($\eta^2(36) = 4.490$, $p < .001$, $d = 1.46$). The results were essentially unchanged when the same analyses were repeated using an arcsine square root transformation to account for inhomogeneity of variance in the percent correct recall data between groups¹.

Table 1 displays the average number of responses generated per trial in each response category, which includes correct responses and incorrect responses, overall and broken down by intrusion type. This table also includes the average “total recall” which is the sum of the average number of correct responses and the average number of intrusions. We include total recall scores because this measure indicates an important group difference that should be considered when interpreting the intrusion rates: OA tended to generate more responses overall than the AD group. That is, in the STM recall task, AD participants recalled fewer studied words ($F(1, 36) = 20.16$, $p < .001$, $\eta^2 = .36$), and made fewer recall responses in total, ($F(1, 36) = 11.59$, $p < .01$, $\eta^2 = .24$), relative to OA participants. The LTM recall results show the same pattern, with AD participants recalling fewer studied words ($F(1, 36) = 89.66$, $p < .001$, $\eta^2 = .71$) and generating fewer responses in total ($F(1, 36) = 68.45$, $p < .001$, $\eta^2 = .66$). In absolute terms, AD participants made significantly more intrusions (of all types) than OA participants in the STM recall task ($F(1, 36) = 8.24$, $p < .01$, $\eta^2 = .186$), but the difference was not reliable in the LTM recall task ($F(1, 36) = 1.813$, $p > .1$, $\eta^2 = .05$). Only a subset of the proportion data showed unequal variances between groups that warranted the use of the arcsine square root transform. Once again, the analyses of the transformed data did not change the results in any conditions².

Note however, that due to differences in the total number of recall responses, direct comparison between groups on absolute number of intrusions is misleading. Therefore,

¹Results from the ANOVA performed on the transformed data are as follows: main effect of group (OA > AD; $F(1,36) = 45.08$, $p < .001$, $\eta^2 = .56$), main effect of task (STM > LTM; $F(1, 36) = 551.97$, $p < .001$, $\eta^2 = .94$), and group $\times$ task interaction ($F(1, 36) = 10.01$, $p < .005$, $\eta^2 = .22$). Post-hoc tests of group differences in LTM ($\eta^2(36) = 8.84$, $p < .001$, $d = 2.87$) and STM ($\eta^2(36) = 4.36$, $p < .001$, $d = 1.4$), were similarly unaffected.

²Results from the analyses of the subset of proportion data that were transformed are as follows: main effect of group for both tasks (STM Recall: $F(1,36) = 12.05$, $p = .001$, $\eta^2 = .25$; LTM Recall: $F(1,36) = 20.13$, $p < .001$, $\eta^2 = .36$), group $\times$ intrusion type interaction for both tasks (STM: $F(2, 35) = 20.1$, $p < .001$, $\eta^2 = .36$; LTM: $F(2, 35) = 8.586$, $p < .001$, $\eta^2 = .19$). The results for proportion of *phonological* intrusions (AD>OA; STM Recall: $\eta^2(36) = 3.99$, $p < .001$, $d = 1.3$; LTM Recall: $\eta^2(36) = 2.474$, $p < .05$, $d = .82$) *other* intrusions (AD>OA; STM Recall: $\eta^2(36) = 4.573$, $p < .001$, $d = 1.5$; LTM Recall: $\eta^2(36) = 4.06$, $p < .001$, $d = 1.3$) and *semantic* intrusions (STM Recall: $\eta^2(36) = -1.48$, $p = .14$, $d = -.48$; LTM Recall: $\eta^2(36) = -1.04$, $p = .30$, $d = -.33$) were similarly unchanged.

intrusion rates were calculated as a proportion of total recall responses (both correct and incorrect; Figure 1) for each task separately. The three intrusion types were semantic, phonological, and other. Using a 2×3 ANOVA (with group as the between-subjects variable and intrusion type as the within-subjects variable) we compared the rate of intrusions (semantic, phonological, and other) between OA and AD participants for each memory task. The results show a main effect of group for both tasks (STM Recall: $F(1,36) = 11.23$, $p < .005$, $\eta^2 = .28$; LTM Recall: $F(1,36) = 14.189$, $p = .001$, $\eta^2 = .28$), reflecting the increased rate of intrusions in AD, and a significant group $\times$ intrusion type interaction for both tasks (STM: $F(2, 35) = 14.82$, $p < .001$, $\eta^2 = .29$; LTM: $F(2, 35) = 9.52$, $p < .001$, $\eta^2 = .21$), indicating that for both tasks, the group difference in intrusion rates depended on the type of intrusion. In both tasks, AD participants had a higher proportion of *phonological* intrusions (STM Recall: $t(36) = 4.12$, $p < .001$, $d = 1.34$; LTM Recall: $t(36) = 2.25$, $p < .05$, $d = .73$) and a higher proportion of *other* intrusions (STM Recall: $t(36) = 3.70$, $p < .01$, $d = 1.20$; LTM Recall: $t(36) = 3.61$, $p < .01$, $d = 1.17$) compared to healthy OAs. The distribution of “Other” intrusions across the subtypes of Unrelated, Non-word and Prior-list Repeats varied by group and task type. In STM unrelated intrusions were most common for both groups. In LTM, of the rare intrusions for the OA group, prior-list repeats were most common, whereas unrelated items were most common for AD. Because of the low values, however, further analyses of the breakdown of “other” intrusions were not conducted. Notably, in neither task did the proportion of *semantic* intrusions differ significantly between groups (STM Recall: $t(36) = -1.78$, $p = .08$, $d = -.58$; LTM Recall: $t(36) = -1.01$, $p = .32$, $d = -.32$). Therefore, AD participants generate significantly more nonsemantic intrusions than healthy OAs, but the rate of semantic intrusions is unaffected by disease status.

Discussion

Participants with AD showed significant impairment in veridical memory when compared to healthy older adults, as measured by lower percent correct recall in both STM and LTM tasks. As expected, AD participants had disproportionately poorer recall in the LTM task. AD participants also produced a higher proportion of intrusions than the OA group in both tasks. The group difference in intrusion rates depended on intrusion type: AD participants generated significantly higher proportions of phonological and other intrusions than OA participants in both tasks. However, in contrast to the other error types, the proportion of semantic intrusions did not differ significantly between the two groups.

Previous DRM studies with AD patients have found an elevated nonsemantic intrusion rate in AD, and have attributed it to a disease-related decline in memory monitoring (Balota et al., 1999, Watson et al., 2001). Similarly, in the current study AD participants committed significantly more phonological and other nonsemantic errors than OA participants, and these types of intrusions made up a larger proportion of their total recall. Importantly, this elevated intrusion rate appears to be independent of veridical recall, as the proportion of phonological and other intrusions remained relatively constant between tasks even as veridical recall decreased (in the STM task, AD participants correctly recalled 65% of the presented items, whereas in the LTM task the percent correct recall dropped to 23%). The proportion of nonsemantic intrusions did not differ between the two tasks in AD ($t(18) = 1.3$, $p > .2$, $d = .24$). Indeed, the comparatively high proportion of phonological intrusions for the AD group in both tasks (.041 and .036 for the STM and LTM tasks, respectively) could be interpreted as evidence for more superficial encoding of the list items by the AD group,⁶ an interpretation that is generally consistent with their more frequent “other” intrusions that included unrelated words. This finding is inconsistent with the report of Watson et al. (2001), who found no group differences in the number of phonological intrusions across three list types (semantic, phonological, and hybrid). However, the authors of this paper

reported intrusions as raw numbers, not as proportions of total recall output, so direct comparison of our results is difficult.

The somewhat indiscriminate nature of the intrusions generated by AD participants in this study indicates a breakdown in their ability to monitor and constrain their responses. Yet the monitoring deficit that is evident in AD does not appear to apply equally to all types of intrusions. In the current study, AD participants produced approximately the *same* proportion of semantic intrusions as control participants in both STM and LTM recall. This finding in LTM is consistent with some prior evidence for false LTM memories but at variance with others (Balota et al., 1999; Watson et al., 2001; Budson et al., 2000; Budson, Sitarski, et al., 2002; Budson, Sullivan, et al., 2002; Gallo et al., 2004; Pierce et al., 2005; Budson et al., 2001; Budson et al., 2003). The reason for these differences among studies is uncertain. Critically, semantic intrusion rates for short-term memory also do not differ between the groups, suggesting that the processes that give rise to these errors may operate at short and long delays (see also Flegal et al., 2010).

In LTM, longer encoding time has been shown to increase false memory errors (McDermott and Watson, 2001). It is not yet known whether this relationship holds for semantic STM errors as well. Due to experimenter error in coding the presentation script, encoding durations were not recorded during the STM task, so encoding times may have differed for the AD and OA groups. Based on the proxy measure of reading speed from the Stroop task, we know that the AD group read significantly fewer words in the timed word reading (without interference) portion of the Stroop task ($t(36) = 3.10, p < .05, d = 1.08$) which suggests that encoding time during the self-paced STM task may also have been slightly longer for the AD group. The potential influence on errors of any encoding time differences that might have existed between the two groups is uncertain however because the relationship of encoding time to false memory rates has yet to be established for STM.

Given the combined evidence from the STM and LTM versions of the task, we are reluctant to conclude that semantic processing is entirely preserved in AD, due to a possible trade-off between monitoring and associative deficiencies. That is, the present evidence for a monitoring deficit would predict that AD participants should produce semantic intrusions at a higher rate than healthy OA participants, assuming that associative semantic processing is spared. The absence of a group difference in semantic intrusion rates could thus result if impaired associative (gist-based) processing in AD is offset by impaired monitoring.

Evidence consistent with this possibility comes from a study by Gallo et al. (2006) in which DRM study items were either blocked by list, or intermixed, which diminishes the strength of the gist memory traces. The authors found that blocking the lists (the high-gist condition) produced a larger increase in false recognition for the control participants than for the AD participants, suggesting that AD patients are impaired in processing gist. True recognition was not affected by blocking for the control group, but AD participants were equally affected by the blocking manipulation in their true and false recognition performance. This suggests that AD participants used gist as their primary source of recognition for both the nonpresented theme words and the studied words, as performance on both judgments suffered in the mixed condition when gist was weakened. The authors concluded that gist is likely impaired in AD, but verbatim memory may be more impaired, and thus patients rely disproportionately on gist when performing a recognition task (see Budson et al., 2003, for a similar conclusion). The results from the current study are consistent with the possibility that gist memory is weakened by AD; however, the resulting proportion of semantic intrusions is unaffected by disease status because AD participants are impaired at distinguishing semantic associates from studied items *and* are less effective at monitoring their responses.

In summary, the present experiment extends prior research demonstrating verbal STM impairment in AD (Belleville et al., 1996; Kensinger, et al., 2003; Peters and Collette et al., 2009; Peters and Majerus et al., 2009) and the pervasive intrusions implicate a role for frontal dysfunction (Budson, Sullivan, et al., 2002; Balota et al., 1999; Mitchell et al., 2006). Moreover, our study provides converging support for two important cognitive effects of AD implicated by prior work, and extends these conclusions by investigating false short-term memory. First, like prior studies (Balota, et al.1999), this study demonstrates that false recall is present as an equivalent proportion of total recall in patients with AD and healthy older adults. We extend this result to demonstrate at least some preservation in AD of associative processing even under STM task conditions, where fewer memory set items and less opportunity for elaboration challenge the formation of such gist-based representations. However, it is possible that the pattern of semantic intrusions seen in both STM and LTM versions of the DRM paradigm can be at least partially attributed to a tradeoff between impaired associative activation/gist processing and impaired monitoring in AD.

Second, as in prior work (Balota et al., 1999; Watson et al., 2001), we demonstrate that patients with AD are more likely to generate nonsemantic intrusions compared to healthy controls. This study offers new evidence of the marked severity of monitoring deficit in AD by revealing that the high proportion of nonsemantic intrusions persists in a short-term memory task where participants can accurately recall almost two-thirds of the memory set. This monitoring deficit is likely due to disease-related frontal lobe pathology (Butler et al., 2004; Budson, Sullivan, et al., 2002; Mitchell et al., 2006) and would be expected to worsen with disease progression (Lowenstein et al., 1991; Rouleau et al., 2001; Fuld et al., 1982). The observation that the proportion of nonsemantic intrusions remains consistent across both STM and LTM conditions despite the large difference in recall accuracy (65% and 23%, respectively) suggests that AD participants are not able to use item-specific recollection to reduce their intrusion rate. These results highlight the fact that AD is characterized not only by memory loss, but also by distortions and confabulations of memory content which likely result from impaired metamemory processes that adversely affect both short and long-term memory performance.

Acknowledgments

We would like to thank Marianne Watson and Halle Zucker for their assistance with participant recruitment, and Dr. Paul Moberg and Hannah McCoubrey for their assistance with neuropsychological testing. This research was supported by NIA Grant R01AG18286 (to P.A.R-L), a Tanner Memorial Research Award to K.M., and NIA Grant 5T32 AG000029-36 (to A.S.A.).

References

- Atkins AS, Reuter-Lorenz PA. False working memories? Semantic distortion in a mere 4 seconds. *Memory and Cognition*. 2008; 36:74–81.
- Baddeley, AD. *Working Memory*. New York: Oxford University Press; 1986.
- Balota DA, Cortese MJ, Duchek JM, Adams D, Roediger HL, McDermott KB, Yearys BE. Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology*. 1999; 16:361–384.
- Belleville S, Peretz I, Melenfant D. Examination of the working memory components in normal aging and in dementia of the Alzheimer type. *Neuropsychologia*. 1996; 34:195–207. [PubMed: 8868277]
- Brainerd, C.J.; Reyna, V.F. *The science of false memory*. New York: Oxford University Press; 2005.
- Budson AE, Daffner KR, Desikan R, Schacter DL. When false recognition is unopposed by true recognition: Gist-based memory distortion in Alzheimer's disease. *Neuropsychology*. 2000; 14:277–287. [PubMed: 10791867]
- Budson AE, Desikan R, Daffner KR, Schacter DL. Perceptual false recognition in Alzheimer's disease. *Neuropsychology*. 2001; 15:230–243. [PubMed: 11324866]

- Budson AE, Sullivan AL, Mayer E, Daffner KR, Black PM, Schacter DL. Suppression of false recognition in Alzheimer's disease and in patients with frontal lobe lesions. *Brain*. 2002; 125:2750–2765. [PubMed: 12429602]
- Budson AE, Sitarski J, Daffner KR, Schacter DL. False recognition of pictures versus words in Alzheimer's disease: the distinctiveness heuristic. *Neuropsychology*. 2002; 16:163–173. [PubMed: 11949708]
- Budson AE, Sullivan AL, Daffner KR, Schacter DL. Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition*. 2003; 51:251–261. [PubMed: 12727179]
- Budson AE, Sullivan AL, Solomon PR, Simons JS, Beier JS, Scinto LF, Daffner KR. Memory and emotions for September 11, 2001, terrorist attacks in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Neuropsychology*. 2004; 18:315–327. [PubMed: 15099154]
- Budson AE, Daffner KR, Dodson CS, Schacter DL. Metacognition and false recognition in Alzheimer's disease: further exploration of the distinctiveness heuristic. *Neuropsychology*. 2005; 19:252–258.
- Butler KM, McDaniel MA, Dornburg CC, Prince AL, Roediger HL. Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review*. 2004; 11:921–925. [PubMed: 15732704]
- Coane JH, McBride DM, Bascom AR III, Jordan JS. False memory in a short-term memory task. *Experimental Psychology*. 2007; 54:62–70. [PubMed: 17341016]
- Dalla Barba G, Nedjam Z, Dubois B. Confabulation, executive functions, and source memory in Alzheimer's disease. *Cognitive Neuropsychology*. 1999; 16:385–398.
- Deese J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*. 1959; 58:17–22. [PubMed: 13664879]
- Folstein MF, Folstein SE, McHugh PR. "Mini mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]
- Flegal KE, Atkins AS, Reuter-Lorenz PA. False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010; 36:1331–1338.
- Fuld PA, Katzman R, Davies P, Terry RD. Intrusions as a sign of Alzheimer dementia: chemical and pathological verification. *Annals of Neurology*. 1982; 11:155–159. [PubMed: 7073248]
- Gallo DA, Daffner KR, Budson AE, Sullivan AL, Schacter DL. Associative recognition in Alzheimer's disease: evidence for impaired recall-to-reject. *Neuropsychology*. 2004; 18:556–563. [PubMed: 15291733]
- Gallo DA, Shahid KR, Olson MA, Solomon TM, Schacter DL, Budson AE. Overdependence on degraded gist memory in Alzheimer's disease. *Neuropsychology*. 2006; 20:625–632. [PubMed: 17100507]
- Gallo DA, Wiseman AL, Chen JM, Schacter DL. Retrieval monitoring and anosognosia in Alzheimer's disease. *Neuropsychology*. 2007; 21:559–568. [PubMed: 17784804]
- Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychological Bulletin*. 1993; 114:3–28. [PubMed: 8346328]
- Jonides J, Lewis RL, Nee DE, Lustig CA, Berman MG, Moore KS. The mind and brain of short-term memory. *Annu Rev Psychol*. 2008; 59:193–224. [PubMed: 17854286]
- Kensinger EA, Shearer DK, Locascio JJ, Growdon JH, Corkin S. Working memory in mild Alzheimer's disease and early Parkinson's disease. *Neuropsychology*. 2003; 17:230–239. [PubMed: 12803428]
- Loewenstein DA, D'Elia L, Guterman A, Eisdorfer C, Wilkie F, LaRue A, Mintzer J, Duara R. The occurrence of different intrusive errors in patients with Alzheimer's disease, multiple cerebral infarctions, and major depression. *Brain and Cognition*. 1991; 16:104–117. [PubMed: 1854466]
- McCabe DP, Presmanes AG, Robertson CL, Smith AD. Item-specific processing reduces false memories. *Psychonomic Bulletin & Review*. 2004; 11:1074–1079. [PubMed: 15875978]
- McDermott KB, Watson JM. The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language*. 2001; 45:160–176.

- McKhann G, Drachman 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; 34:939–944. [PubMed: 6610841]
- Mitchell JP, Schacter DL, Sullivan AL, Budson AE. Misattribution errors in Alzheimer's disease: The illusory truth effect. *Neuropsychology*. 2006; 20:185–192. [PubMed: 16594779]
- Nelson, DL.; McEvoy, CL.; Schreiber, TA. The University of South Florida word association, rhyme, and word fragment norms. 1998. Retrieved from <http://www.usf.edu/FreeAssociation/>
- Peters F, Collette F, Degueldre C, Sterpenich V, Majerus S, Salmon E. The neural correlates of verbal short-term memory in Alzheimer's disease: an fMRI study. *Brain*. 2009; 132:1833–1846. [PubMed: 19433442]
- Peters F, Majerus S, De Baeremaeker J, Salmon E, Collette F. Impaired semantic knowledge underlies the reduced verbal short-term storage capacity in Alzheimer's disease. *Neuropsychologia*. 2009; 47:3067–3073. [PubMed: 19607851]
- Pierce BH, Sullivan AL, Schacter DL, Budson AE. Comparing source-based and gist-based false recognition in aging and Alzheimer's disease. *Neuropsychology*. 2005; 19:411–419. [PubMed: 16060815]
- Reuter-Lorenz, PA.; Jonides, J. The executive is central to working memory: Insights from age, performance and task variations. Conway, RA.; Jarrold, C.; Kane, MJ.; Miyake, A.; Towse, J., editors. *Variations in Working Memory*: Oxford University Press; 2007. p. 250-270.
- Robinson KJ, Roediger HL III. Associative processes in false recall and false recognition. *Psychological Science*. 1997; 8:231–237.
- Roediger HL III, McDermott KB. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1995; 21:803–814.
- Roediger HL III, Watson JM, McDermott KB, Gallo DA. Factors that determine false recall: a multiple regression analysis. *Psychonomic Bulletin & Review*. 2001; 8:385–407. [PubMed: 11700893]
- Rouleau I, Imbault H, Laframboise M, Bédard MA. Patterns of intrusions in verbal recall: Comparison of Alzheimer's disease, Parkinson's disease, and frontal lobe dementia. *Brain and Cognition*. 2001; 46:244–249. [PubMed: 11527341]
- Schacter DL, Israel L, Racine C. Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*. 1999; 40:1–24.
- Watson JM, Balota DA, Sergent-Marshall SD. Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*. 2001; 15:254–267. [PubMed: 11324868]

A. STM Recall: Proportion of Total Recall

B. LTM Recall: Proportion of Total Recall

Figure 1.

Mean proportion of total recall for A) short-term memory (STM) recall task and B) long-term memory (LTM) recall task. Error bars indicate standard error of the mean. OA: Older Adult participants, AD: Alzheimer's Disease participants.

Table 1

Mean Number of Recall Responses per List as a Function of Group and Response Type

A. Short-Term Memory Recall (from 4-word lists)

	Older Adult		Alzheimer's Disease	
	<i>Mean</i>	<i>(SD)</i>	<i>Mean</i>	<i>(SD)</i>
Total Recall	3.59	(.25)	3.17	(.48)
Correct Recall	3.37	(.28)	2.61	(.67)
Total Intrusions	.22	(.18)	.56	(.47)
Semantic Intrusions	.15	(.11)	.08	(.06)
Phonological Intrusions	.02	(.04)	.13	(.11)
Other Intrusions	.05	(.07)	.35	(.37)

B. Long-Term Memory Recall (from 12-word lists)

	Older Adult		Alzheimer's Disease	
	<i>Mean</i>	<i>(SD)</i>	<i>Mean</i>	<i>(SD)</i>
Total Recall	7.10	(1.0)	3.68	(1.5)
Correct Recall	6.42	(1.0)	2.76	(1.4)
Total Intrusions	.68	(.40)	.92	(.69)
Semantic Intrusions	.60	(.37)	.25	(.27)
Phonological Intrusions	.03	(.06)	.13	(.22)
Other Intrusions	.05	(.10)	.54	(.60)