DIFFERENTIAL RESPONSE BIASES BETWEEN OLDER AND YOUNGER ADULTS ON RECALL TESTS

by

Mark Jordan Huff

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Psychological Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2010

APPROVAL

of a thesis submitted by

Mark Jordan Huff

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Michelle L. Meade, Ph.D.

Approved for the Department of Psychology

Ruth Striegel-Moore, Ph.D.

Approved for the Division of Graduate Education

Carl A. Fox, Ph.D.

STATEMENT OF PERMISSION TO USE

Mark Jordan Huff

April 2010

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

This study examined the role of differential response biases between older and younger adults on recall performance and confidence judgments. Participants studied categorized and unrelated word lists and were asked to recall the items under one of the following test conditions: standard free recall, free recall with a penalty for guessing, free recall with no penalty for guessing, or forced recall. The results showed that both older and younger adults adjusted their responses in accordance with test instructions. Older adults reported a greater number of intrusions relative to younger adults in both standard and penalty instructions, particularly those intrusions that were plausible given a prior study list. Additionally, older adults were more likely to report intrusions with greater confidence that they were presented on an earlier studied list. Results demonstrate that older adults report memory information on free recall tasks with a liberal response bias, as evidenced by a decreased ability to withhold reporting intrusions and claiming to remember intrusions with greater confidence compared to younger adults.

INTRODUCTION

There is substantial evidence that memory abilities of older adults suffer relative to younger adults (see Balota, Dolan, & Duchek, 2000, for a review); however, relatively little research has examined the role of a response bias on older adults' memory performance. Age-related response biases on recall tests refer to the idea that older and younger adults might differ in the amount of memory information needed before they are willing to report a memory response. In other words, older and younger adults might rely on different information to make responses on a memory test. Recent evidence supports the idea of an age-related response bias on various memory tasks (e.g., Jacoby, Hessels, & Toth, 2005; Kelley & Sahakyan, 2003; Pansky, Goldsmith, Koriat, & Pearlman-Avnion, 2009), but the current study is the first to examine potential age-related response biases on free recall. Specifically, the current study further explores possible age-related response biases by examining age differences in free and forced recall tests using different instructional manipulations (e.g., penalties against guessing) and varying stimulus materials (unrelated and categorized word lists).

Possible response biases on free recall tests have been examined by comparing differential memory responding on tests using free or forced recall instructions (Cofer, 1967; Erdelyi, Finks, & Feigin-Pfau, 1989; McKelvie, 2001; Ritter & Buschke, 1974; Roediger & Payne, 1985). Free recall instructions allow participants to determine their own criterion for reporting and withholding items on a recall test, while forced recall requires that participants produce a given number of responses, even if guessing is required. The rationale for requiring participants to report guesses on a recall test was based on the assumptions of generate-recognize models of recall (e.g., Anderson & Bower, 1972; Anderson, 1983) which hypothesized that participants complete a recall test by generating possible items and then selecting a subset of the generated items to report. Forced recall procedures were adopted to equate response criteria between participants and ultimately determine whether free recall tasks are conservatively biased (i.e., participants can recall more items than they report on the recall test). Much research comparing free and forced recall tests suggests that in fact, recall is not a conservatively biased test, as the number of correct items produced on free and forced recall tasks are generally equivalent (see Roediger, Wheeler, & Rajaram, 1993, for a review). Participants produce more items overall on forced recall tests, consistent with task demands, but the additional items produced are intrusions, rather than additional correct items. One important exception to this conclusion is that forced recall does show slight increases in correct items over free recall when the test materials lend themselves to guessing (e.g. categorized pictures and concrete nouns, Erdelyi et al., 1989; Roediger, Srinivas, & Waddill, 1989). Importantly, this pattern also applies to older adults. Forced recall improves older adults' memory relative to cued recall for categorized lists, though older adults report fewer correct items on an initial forced recall test than younger adults (Meade & Roediger, 2006; see too Meade & Roediger, 2009). However, for unrelated materials, older and younger adults both derive no benefit from forced recall relative to free recall (Henkel, 2007).

In the current study, we further explore the role of response bias on free recall tasks, but rather than asking if recall is a conservatively biased test, we focus instead on possible age differences in how participants might approach free and forced recall tests. Do younger and older adults differ in the amount of information required to report a memory response? Do potential age-related response biases vary in relation to instructional manipulations (standard free recall, penalties against guessing, encouragement to guess, and forced recall)? To answer such questions, we examined both veridical and false recall (intrusions) as well as participants' confidence that a reported memory item was correct. Intrusions may be especially important in regards to response biases. Given the research findings that forced recall does not generally increase the number of correct items relative to a free recall test, individuals who are biased to respond liberally will most likely increase their response output by increasing the number of incorrect items. Oppositely, if individuals are biased to respond conservatively, they may withhold items for which they do not have enough memory evidence, and minimize the number of intrusions. Confidence ratings are also quite important to potential response biases because a liberal response bias might also allow for the reporting of memory items that are of lower confidence compared to a conservative response bias in which only responses with the highest confidence will be reported. If older and younger adults recall memory items differently, they are likely to differ in the number of intrusions reported and the confidence with which they do so. Finally, consistent with early research on free and forced recall, we include both categorized word lists and unrelated word lists so as to vary how likely guessing is to increase recall.

Based on research conducted in other paradigms, there is growing evidence to suggest that older adults do indeed rely on a different basis for responding than do young adults. Age-related response biases have been explored on memory tasks that utilize the metacognitive model proposed by Koriat and Goldsmith (1996). Specifically, the regulation of memory responses between accuracy and quantity has been described in terms of strategic ability. Quantity is an input bound measure in which the assessment is based on all of the memory items that are presented at study. Accuracy is an output-bound measure where an individual has control over the items reported and the willingness to report an item from a memory set is contingent in one's subjective confidence that the item was studied earlier. Accuracy and quantity are described in terms of a tradeoff where individuals can maximize the accuracy of their memory report but at the expense of quantity or vice versa. The quantity/accuracy tradeoff can only be assessed under conditions of free report where individuals have the opportunity to volunteer or abstain from reporting information. The quantity/accuracy tradeoff that is found under free-report conditions, suggests the use of a free-report recall task is essential for determining a response bias and the number of intrusions an individual is willing to report. Similarly, Koriat and Goldsmith suggest that individuals monitor their memory responses and terminate a memory search when an individual judges a "set of accessible memory items to be depleted and more effort would be futile" (pp. 490). Therefore, the critical difference between individuals, and perhaps between age groups, is when a memory search is judged as no longer beneficial to memory performance.

According to the Koriat and Goldsmith (1996) metamemory model, memory performance is contingent upon three factors. The first factor is an individual's monitoring effectiveness, which is the ability one has to differentiate between a correct or incorrect response from within a studied memory set. The second factor is control sensitivity in which the rememberer bases their decision to report a response based on the responses' assessed probability of being correct. Control sensitivity is contingent upon an individuals' motivation to perform accurately or based on quantity, which is dependent upon external incentive. The third factor is the response criterion setting which an individual can adjust based on consequences for commission errors. Assessment of these three factors requires participants to first provide a memory under forced-report conditions followed by a subsequent free-report test where the individual has the option to withhold a memory response they are not confident is correct. The comparison of forced and free-report conditions allows for the assessment of the willingness for an individual to report from memory.

More recently, researchers have used the metamemory framework designed by Koriat and Goldsmith (1996) to examine the differences in responding between older and younger adults. Kelly and Sahakyan (2003) evaluated younger and older adults on their abilities to strategically regulate their memory accuracy. In their design, younger and older adults were presented with word pairs that were unrelated (clock – dollar) or deceptive (nurse – dollar) during the study phase. Notice that the deceptive pairs have one word that could be replaced by a semantically similar word (e.g. doctor) that shares a similar phonological letter structure. After the study trial, participants completed a forced-report cued recall task where one cue was presented and the participant was instructed to provide the other item that was presented simultaneously at study and estimate the likelihood that the response reported was correct. Participants then completed a subsequent free-report cued recall test where the same cue was presented, but participants had the option to withhold their response, similar to Koriat and Goldsmith's paradigm. Additionally, free-report conditions were completed under various incentives designed to increase the accuracy of participants responding. Results indicated that older adults were less accurate in their memory responses and this decrease in accuracy was even greater for the plausible deceptive items. Incentives increased the accuracy of younger adults but did not influence the accuracy of older adults. Critically, when younger and older adults were compared on their monitoring effectiveness, the correspondence between confidence assessments and correct responding, both groups were equal on control items, but older adults were decreased in their monitoring effectiveness for deceptive items, reporting deceptive items with greater confidence. Differences in monitoring effectiveness between younger and older adults suggest a different basis for responding from memory. The use of deceptive information appears to inflate the differences between younger and older adults in their ability to make an accuracy assessment. Older adults also appear to respond from memory with items that appear plausible.

In a similar study using the same memory monitoring framework of Koriat and Goldsmith (1996), Pansky et al. (2009) provided older and younger participants with pictorial scenes for study followed by a forced-report multiple-choice test. During the presentation of the scenes, younger adults encoded stimulus materials under full or divided attention. The encoding manipulation was designed to equate the quantitative memory performance between younger and older adults. Participants were required to provide an estimate of their confidence that their response was correct prior to completing a free-report phase where their responses could be withheld. Results demonstrated that older adults were more likely to volunteer responses, both correct and incorrect, than younger adults, even when younger adults' attention was divided at study. Pansky et al. interpreted these findings as evidence that older adults have a more liberal basis for responding on memory tests. The greater number of reported intrusions for older adults, even when encoding was equated between age groups, suggests that older adults possess a different report criterion and are more liberal in their willingness to report a response compared to younger adults.

In a related study, Jacoby, Bishara, Hessels, and Toth (2005) presented participants with valid (correct), invalid (misleading), or baseline (ampersands) prime-target pairs. The invalid pairs were designed to be misleading by capturing the responses of participants such that, on an initial forced-report cued recall test, they would respond to primes with a plausible, but incorrect target. When given the opportunity to improve their response accuracy on the subsequent free-report cued recall phase, younger adults were able to increase their accuracy by 30%; however, older adults only increased their reporting accuracy 3%. This suggests that older adults are less able to monitor the accuracy of the items produced, and therefore report a higher rate of errors. Younger adults on the other hand, appear to be better able to monitor their own accuracy, even in the face of deceptive, misleading items.

Considered together, these studies provide converging evidence that older adults may have a different basis for responding on memory tests than do younger adults and further specify that older and younger adults not only differ in their quantitative output from memory, but also the accuracy with which they respond. Specifically, older adults volunteer a greater number of responses than do younger adults (Pansky et al., 2009), they rely disproportionately on guesses and/or plausibility (Kelley & Sahakyan, 2003), and they are less able to improve memory accuracy by withholding items (Jacoby et al., 2005). Of interest to the current study is whether the age-related changes in response biases obtained in other paradigms will also apply to performance on free recall and forced recall tests, thus adding generality to previous findings.

The present study assesses response biases by examining performance differences between age groups using a single free recall task. Free recall is unique relative to other memory assessments in that it relies upon an individuals' self-initiated retrieval of a memory episode with the option to withhold a response (Craik & McDowd, 1987). Additionally, free recall is an unbiased memory process because it does not provide additional contextual information present in other tasks such as recognition or cued recall and it does not encourage participants to guess or withhold responses under standard instructions. In regards to aging, differences between older and younger adults appear to be most prevalent in free recall performance compared to other memory tasks, indicating that free recall performance is especially sensitive to aging differences (Anderson, Craik, & Naveh-Benjamin, 1998; Park, Smith, Dudley, & Lafronza, 1989; Veiel & Storandt, 2003). Free recall performance also appears to correlate well with one's age in longitudinal and cross-sequential studies demonstrating a linear decline in free recall performance overtime (Park et al., 1996; Ronnlund, Nyberg, Backman, & Nilsson, 2003).

In essence, if a response bias occurs between older and younger adults, a free recall task would likely demonstrate the differences through the number of reported intrusions due to its sensitivity to age-related memory decline.

Further, the current paradigm includes several changes in methodology relative to previous research that may further inform the nature of age-related biases on recall. First of all, the current experiment utilizes a single recall test, rather than the repeated tests utilized in previous literature. It is difficult to assess a participants' natural memory response when they are tested in forced-report followed by free-report. Using a single free recall task, a participant has the option to withhold a response initially without the influence of a prior test. Additionally, the present design compares response differences using thematically categorized and unrelated word lists compared to constrained cues which may artificially create memory errors. The materials in the present study are not designed to be misleading, but allow for a plausible response when reporting from a category. Finally, the present study uses several different instructions that are designed to increase or decrease an individuals' willingness to report from memory (see too Kelley & Sahakyan, 2003). For example, if participants are encouraged to guess, they are likely to generate responses in which they may not have great confidence and would otherwise have withheld. By varying recall conditions using penalties against guessing, encouraging guessing, or forced recall, older and younger adults may differentially regulate their own willingness to report intrusions.

The current experiment was designed to examine potential differences in response bias between older and younger adults on a free recall task. Participants were presented with both categorized and unrelated word lists and were given one of four different instructions at recall which were designed to adjust the number of intrusions due to guessing. Standard free recall instructions were used to assess the natural responding of participants without the suggestion of guessing or a penalty associated with an incorrect response. Penalty instructions were used to examine the ability of participants to eliminate erroneous responses from being reported due to a penalty assigned to any incorrect response. Guessing instructions, on the other hand, assessed participants' willingness to report intrusions by encouraging them to guess and report items in which they have little or no confidence. Participants in forced instructions were required to match their recall to the total number of items presented at study. Under forced instructions, it was assumed that participants would not be able to recall all of the memory items studied and would therefore be required to guess. Across all instructions, participants were required to provide an estimate of their subjective confidence that the item they reported was correct.

For correct recall, it was expected that both age groups would be able to successfully alter their correct recall levels in accordance with test instructions, but only after studying a categorized word list. Specifically, we predicted that participants would be able to increase their correct recall under guessing and forced recall conditions, relative to standard instructions, because participants could provide categorically consistent responses (Erdelyi et al., 1989). Under penalty instructions, participants were expected to decrease the total number of correct items, relative to standard instructions, by withholding low confidence responses. We also do not predict differences between age groups on categorized word lists in the various instructions because correct recall rates are typically matched between age groups on categorical word lists (Luszcz, Roberts, & Mattiske, 1990). Instructional differences were not expected on unrelated word lists due to the lack of thematic consistency and the seemingly random relationship between word items. Additionally, no differences between older and younger adults were predicted based on instructions on unrelated word lists because guessing low confidence responses would not benefit recall. As mentioned above, we do not expect differences in correct recall of categorized and unrelated lists to be attributable to a response bias due to the relatively little gain in correct recall compared to the large increase in intrusions when comparing forced and free recall. Rather, a response bias is likely to operate on the number of intrusions reported.

It was expected that intrusions would vary based on test instructions. For related lists, intrusions were expected to increase in more liberal conditions (i.e. guessing and forced recall conditions), but would decrease in the penalty condition relative to standard instructions. However, unlike correct recall, intrusions were expected to vary to a larger extent than correct recall based on instructions with large increases in the number of intrusions reported per list when participant were encouraged to guess or forced to guess, but also substantial decreases with more conservative instructions. Age differences were predicted to impact the willingness to report intrusions such that older adults would be biased to respond liberally relative to younger adults, reporting a greater number of intrusions per list under standard instructions (cf. Pansky et al., 2009). This response bias will also impact instructional conditions where we predict that although older adults will report a greater number of intrusions under standard instructions, younger adults will be more likely to take advantage of test instructions by increasing the total number of intrusions when encouraged to do so, and minimizing the number of intrusions when a penalty is in effect. On categorized lists, older and younger adults are also expected to alter the number of intrusions based on task demands, but older adults are predicted to report more intrusions across instructions. This is because the thematic focus of a categorized list allows for the production of intrusions that are plausible. We predict that older adults in particular will report intrusions due to their tendency to report from memory based on plausibility (Jacoby et al., 2005; Kelley & Sahakyan, 2003).

For unrelated lists, intrusions are likely to be minimal due to the lack of thematic consistency between items. We predict that participants will increase or decrease the number of intrusions based on instructions; however, older adults are predicted to report a greater number of intrusions compared to younger adults, particularly under conservative recall instructions. This is because older adults are not able to monitor the correctness of their responses to the same extent as younger adults (Jacoby et al., 2005; Kelley & Sahakyan, 2003).

Confidence ratings are also expected to provide information on the types of intrusions reported by age groups. Specifically, we predict that older adults are likely to report intrusions with greater confidence than younger adults. For plausible intrusions that are thematically consistent with a studied list, we predict that older adults may be especially overconfident that plausible intrusions are correct (Kelley & Sahakyan, 2003). Additionally, test instructions may also influence the confidence with which participants report from memory. For instance, when completing a forced recall test, as memory items become exhausted, participants will be required to guess, resulting in a decrease in confidence. When test instructions result in a penalty for guessing, participants may only report items in which they have high confidence and withhold any low confidence items. Taken together, test instructions will likely influence the confidence with which memory items are reported and the overconfidence elicited by older adults is predicted to pervade across test instructions.

METHOD

Participants

A total of 80 undergraduates at Montana State University (age range of 17 – 39 years) and 80 community dwelling older adults (age range of 63 – 89 years) were recruited for participation in the study. The younger adults participated for course credit as part of an introductory Psychology class, while older adults were given a $10 compensation for their participation in the study. Older adults were recruited through newspaper advertisements and from the MSU older adult subject pool from the Bozeman community. Older Adults were responsible for transportation to and from the testing site at the university and both younger and older adults completed the experiment at the same site.

To assess overall cognitive performance, we administered the Mini Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) in addition to the Shipley Institute of Living Scale (multiple-choice vocabulary test; Shipley, 1940) at the end of the experiment. Participants who scored within the normal range on the MMSE (score ≥ 26 out of 30) qualified their data for analysis in the current study. Table 1 shows the scores of the above measures as well as demographic information for both younger and older adults. Younger adults (M = 29.20) scored significantly higher on the MMSE than older adults (M = 28.70), $t(158) = 2.88$, $SEM = 0.10$, $p = .005$. These differences were not a concern because all participants scored high enough to place within the normal healthy range for cognitive performance. Shipley vocabulary scores were also compared between age groups, with older adults (M = 34.73) significantly outperforming younger adults (M

= 28.07, $t(152) = 11.84$, $SEM = 0.42$, $p < .001$). Older adults ($M = 16.06$) had also attained a greater number of years of education than younger adults ($M = 13.16$, $t(158) = 9.23$, $SEM = 0.13$, $p < .001$). This pattern was not surprising given that younger adult participants were enrolled in a freshman-level course. Additionally, younger adult participants ($M = 19.90$) were significantly younger than older adult participants ($M = 73.29$, $t(158) = 69.73$, $SEM = 0.34$, $p < .001$).

Table 1. Mean Scores on the Mini Mental State Exam (MMSE) and Shipley Vocabulary Test, Mean Age (yrs), and Education (yrs), and Mean Score on the Memory Anxiety Questionnaire (Scale 1-5) for Younger and Older Adults.

	Younger (N = 80)	Older (N = 80)
MMSE	29.20	28.70
Shipley	28.07	34.73
Age	19.90	73.29
Education	13.16	16.06
Memory Anxiety	2.95	3.04

Design and Materials

The design was a 2 (age group: younger vs. older adult) × 4 (test instructions: standard vs. penalty vs. no penalty vs. forced) × 2 (list type: unrelated vs. related) mixed design. Age and instructional conditions were between-subjects manipulations and all participants received the related and unrelated word lists.

Study Lists

The Battig and Montague (1969) categorical word association norms were used to construct four word lists of 20 items that were all related by one overlying categorical theme and four lists of 20 items that were all unrelated. The related lists were constructed by using the top 25 exemplars from four different categorical lists; each word was matched to an unrelated word that had the same frequency of occurrence in the English language. The unrelated words were selected from the English Lexicon Project database (Balota et al., 2007). The unrelated words were also matched to the related words on the basis of concreteness (being an actual item), familiarity (knowledge or awareness of an item), and imageability (the ease to which the item comes to mind) using The University of Western Australia MRC Psycholinguistic Database (Coltheart, 1981).

Guess-Check Procedure

The matched materials yielded a total of eight 25-item word lists. Each of these word lists were then divided into two versions. The top 10 exemplars in each list were divided into two five-word segments with the last 15 items remaining the same. The formation of two versions allowed for the opportunity to check for guessing since five of the most commonly generated categorical items were not presented during the study phase [see Meade and Roediger (2009) for a similar procedure using a guess-check]. The study lists used in the current study, along with their division into separate versions, are presented in Appendix A.

Memory Anxiety Questionnaire

The Memory Anxiety Questionnaire (adopted from Davidson, Dixon, & Hultsch, 1991), assessed participants' anxiety about their memory performance on a scale from 1 to 5. Higher ratings on the questionnaire indicated a greater level of memory anxiety. Memory anxiety ratings are reported in Table 1. A 2 (age group: old vs. young) × 4 (test instructions: standard vs. penalty vs. no penalty vs. forced) between-subjects analysis of variance (ANOVA) was completed on anxiety scores. No main effects were found for age group or test instructions (both *F*s < 1.86, *p*s $> .14$) and the interaction between age groups and test instructions failed to reach significance, $F(3, 149) = .79$, $MSE = 0.40$, $p > .05$. Anxiety scores from three younger adults were missing due to experimenter error; this was not considered a concern due to the large number of participants in each condition. The lack of effects for memory anxiety suggests that any results obtained in the current study are not due to differences between younger and older adults' concern with memory performance. Correlations between the anxiety questionnaire and memory performance measures are further examined below.

Additional Questionnaires

Additional materials included a locally developed demographics questionnaire, the MMSE, the Shipley Vocabulary Test, and the Memory Anxiety Questionnaire. The memory anxiety questionnaire was taken from Davidson et al., (1991) and modified to include 12 statements and a 5-point likert scale to assess varying levels of agreement with each statement. The modified Memory Anxiety Questionnaire used in the current study is presented in Appendix B.

Procedure

Participants completed the experiment individually or in groups of two without collaboration. After the completion of the consent form, participants were presented with 8 study-test trials. During study, participants were informed to pay attention to each of the word items presented on the computer screen in preparation for a memory test. The words were presented in a black 24-point font on a white background. Presentation of materials was identical for all conditions and age groups. Following the word list presentation, participants completed a 90 second math filler task. During the test, participants were provided with a recall sheet to complete a recall task. All participants were informed that points would be awarded for correct responses and those who perform well on the recall tasks would be entered into a raffle drawing for a prize. These instructions were given to equate motivation across all conditions. Participants were then given one of four different types of recall instructions intended to vary the response criterion setting. The Standard Free Recall conditions informed participants to write down only words they remembered as being presented on the computer screen. In the Penalty condition, participants were informed that their score on the test would be based on the correct responses with the subtraction of any incorrect responses, and therefore, it would be advantageous to write those items they were absolutely positive occurred on the study list. In the No Penalty condition, participants were told that no deduction was associated with an incorrect response, and reporting low confidence responses would be beneficial to their overall performance. The final condition included Forced Recall instructions in which participants were required to provide a response for every blank space on their recall sheet, providing a total of 20 responses. The forced recall condition was designed to elicit any response that may come to mind during the recall phase. Additionally, while recalling items, participants in every condition were required to provide a confidence rating based on their recorded response. Participants indicated confidence using a four-point scale (1-4), with a 4 rating made if the response was provided with highest confidence, a 3 rating if the participant was somewhat confident, a 2 rating for somewhat not confident, and a 1 rating if the participant had very little or no confidence in his or her reported response. Participants were given three minutes to complete the recall task, with more time provided if needed. Participants rarely used more than three minutes for recall.

Participants completed the above procedure for each of the eight word lists, which were presented in a random order. Retrieval instructions were repeated prior to each recall phase to increase instructional salience to the participant. Following memory testing, all participants completed several questionnaires which included demographics information, MMSE, Shipley Vocabulary, and the Memory Anxiety Questionnaire. Following the assessments, participants were fully debriefed, and older adults were provided with compensation. A typical experimental session lasted slightly over an hour.

RESULTS

Correct Recall

The proportion of correctly recalled items along with the proportion of items participants assigned the highest confidence rating of 4 are reported in Table 2. A 2 (Age Group: young vs. old) × 2 (List Type: related vs. unrelated) × 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) mixed factorial ANOVA, with list type as a within-subjects variable, was conducted on the proportion of correctly recalled items. A significant main effect of age was found, $F(1, 152) = 56.37$, $MSE = 0.01$, $p < .001$, suggesting that across related and unrelated lists, older adults ($M = .36$) recalled fewer items than younger adults ($M = .45$). A main effect of list type, $F(1,152) = 2918.11$, $MSE = 0.003$, $p < .001$, demonstrates a greater proportion of correctly recalled items on the related word lists compared to unrelated word lists. Additionally, the main effects of age and list type were qualified by a significant List Type × Age Group interaction, $F(1, 152) = 26.41$, $MSE = 0.003$, $p < .001$. Follow up tests confirmed that although older adults recalled a smaller proportion of items than younger adults on both unrelated word lists ($M = .18$ for older adults, $M = .30$ for younger adults, $t(158) = 9.33$, $SEM = 0.01$, $p < .001$), as well as related word lists ($M = .54$ for older adults, $M = .60$ for younger adults, $t(158) = 4.10$, $SEM = 0.01$, $p < .001$), the age difference was larger for unrelated lists than for lists that offered categorical organization as retrieval support (cf. Rabinowitz, Craik, & Ackerman, 1982).

The main effect of test instructions was only marginally significant, *F*(3, 152) = 2.37, *MSE* = 0.03, p = .07, but the interaction between list type and test instructions was significant, *F*(3, 152) = 6.64, *MSE* =0.02, $p < .001$. Follow up comparisons indicated the proportion of correctly recalled items varied in accordance with test instructions, but only for related word lists (recall on the unrelated lists did not vary as a function of test instructions, *t*s < .76, *p*s > .05). Specifically, for related lists, participants in the forced recall conditions (*M* = .60) were able to increase their veridical recall relative to penalty (*M* = .53, *t*(78) = 3.51, *SEM* = 0.01, p = .001), and standard instructions (*M* = .55, *t*(78) = 2.13, *SEM* = 0.01, $p < .04$), a finding consistent with previous work suggesting that forced recall may sometimes improve veridical recall for categorized list materials (Meade & Roediger, 2006). Participants were also able to increase veridical recall under the no penalty instructions (*M* = .59) relative to the penalty instructions (*M* = .53, *t*(78) = 3.15, *SEM* = 0.01, $p < .01$). All other differences in test instructions were not significant (*t*s < 1.67, *p*s > .10). Finally, the three-way interaction between age, list type, and test instructions failed to reach significance, $F < 1$. Considered together, correct recall data suggests that although the proportion of items correctly recalled could vary based on test instructions when word items were related, older and younger adults were able to adjust this proportion similarly, demonstrating that age groups were not biased to respond differently based on the proportion of items correctly recalled.

Table 2. Mean proportion of items correctly recalled by younger and older adults on recall tests with standard instructions, penalty for guessing instructions, no penalty for guessing instructions, and forced recall instructions (in bold) along with proportion of responses assigned confidence rating of 4 (highest confidence in correct response).

	Younger (N = 80)		Older (N = 80)	
	Related	Unrelated	Related	Unrelated
Standard Total	**.59**	**.31**	**.52**	**.18**
CR 4	.54	.29	.45	.17
Penalty Total	**.55**	**.28**	**.50**	**.18**
CR 4	.50	.27	.44	.17
No Penalty	**.63**	**.30**	**.55**	**.17**
CR 4	.57	.28	.45	.14
Forced	**.61**	**.30**	**.59**	**.18**
CR 4	.53	.29	.50	.16

Confidence Ratings (Correct Recall)

Confidence ratings were collected to assess a participants' belief that a reported response was correct. Confidence ratings may be especially important to age differences in recall across test instructions, because, relative to young adults, older adults demonstrate a lower correspondence between confidence ratings and accuracy (Kelley & Sahakyan, 2003; Pansky et al., 2009). Therefore, differences in participants' subjective judgment in the correctness of their responses may differ between age groups or recall instructions even if there is little or no variance in the proportion of items correctly recalled.

The proportion of correct responses made with a confidence rating of 4 (highest confidence the response was correct) was analyzed in a 2 (Age Group: young vs. old) × 2 (List Type: related vs. unrelated) × 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) mixed factorial analysis of variance (ANOVA) with list type as a within-subjects variable. As is evident in Table 2, the patterns of results for highest confidence ratings closely resemble the results reported for the proportion of correct responses. Statistically, all main effects and interactions were replicated with the exception that for confidence ratings, there was a marginal List Type × Test Instructions interaction, $F(3, 152) = 2.23$, $MSE = 0.003$, $p = .09$. However the trend was similar to that of the proportion of correct recall.

First Ten Items

Guessing is critically important to our research questions, because to the extent that participants are changing their criterion across test conditions, we might expect to see the most flexibility among items from categories that are easily guessable. To examine guesses on the memory test, we incorporated a guess-check procedure which allowed for the assessment of participants guessing plausible nonpresented items (see Meade & Roediger, 2009). As explained above, the top 10 exemplars in the related and unrelated word lists were factored out of the study lists and split into two groups of five items each. Five items were presented to participants, and five items were designated as critical nonpresented items. Across participants, we could assess how likely they were to recall a given item when it was studied relative to how likely they were to guess that same item simply because it was a typical exemplar of a presented category. This procedure is especially important on related word lists because the first items in the norms are the most typical exemplars of a given category, and participants may produce them by generating guesses rather than by retrieving them from memory. Additionally, the guess-check procedure may be especially sensitive to the guessing of older adults as they tend to erroneously report plausible memory items (Kelley & Sahakyan, 2003). If age differences in memory responding occur, they are likely found in intrusion rates, particularly those that are thematically similar to a previous memory set.

First Ten Items (5 Presented)

The mean proportions of the first five items correctly recalled are presented at the top of Table 3. A 2 (Age Group: old vs. young) × 2 (List Type: related vs. unrelated) × 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) mixed factorial ANOVA with list type as a within-subjects variable was computed on the mean proportion of the five presented items. This demonstrated a main effect of age group, $F(1, 152) = 42.93$, $MSE = 0.02$, $p < .001$, suggesting that across related and unrelated lists, younger adults ($M = .58$) recalled more of the first five items than older adults ($M = .47$). A main effect of list type, $F(1, 152) = 721.91$, $MSE = 0.01$, $p < .001$, indicated that across recall instructions and age groups, the five presented items were more likely to be recalled on related lists ($M = .69$) than unrelated lists ($M = .37$). Additionally, the main effects of age and list type were qualified by a significant List Type × Age Group interaction, $F(3, 152) = 43.65$, $MSE = 0.01$, $p < .001$. Follow up tests demonstrated that older adults recalled fewer items than younger adults on unrelated word lists (*M* = .27 for older adults, *M* = .46 for younger adults, $t(158) = 8.76$, *SEM* = 0.02, $p < .001$), but both age groups recalled an equivalent number of items on related word lists (*M* = .67 for older adults, *M* = .70 for younger adults, $t(158) = 1.51$, *SEM* = 0.02, $p = .13$), thus the age difference for correct recall for the first five presented items was driven primarily by memory performance differences in unrelated word lists.

A significant main effect of test instructions, $F(3, 152) = 3.48$, *MSE* = 0.02, $p < .02$, demonstrated that participants adjusted the proportion of recall for the five presented items depending on recall instructions. This main effect was further qualified by a significant List Type × Test Instructions interaction, $F(3, 152) = 8.22$, *MSE* = 0.01, $p < .01$. Follow up comparisons indicated that test instructions did not influence recall of the five presented items for unrelated word lists (all *t*s < 1.65, *p*s > .11), but that test instructions did influence the proportion of correct recall on related word lists. Specifically, post hoc *t*-tests revealed that the proportion of correct recall increased in the forced recall condition (*M* =.78) relative to standard (*M* = .67, $t(78) = 4.00$, *SEM* = 0.02, $p < .001$), penalty (*M* = .65, $t(78) = 4.63$, *SEM* = 0.02, $p < .001$) and no penalty conditions (*M* =.65, $t(78) = 5.30$, *SEM* = 0.02, $p < .001$). Additionally, this pattern of results did not differ based on age group as the three-way interaction between list type, test instructions, and age group, failed to reach significance, $F(3, 152) = 1.42$, $p > .05$. These data replicate past literature comparing forced and free recall in which forced instructions increase recall relative to standard free instructions, but only when guessing could yield a correct response (in this case a categorized word list, Erdelyi et al., 1989; Roediger & Payne, 1985).

Table 3. Mean proportion of first ten items (five presented, five critical nonpresented, and corrected items) recalled by younger and older adults on recall tests with standard instructions, penalty for guessing instructions, no penalty for guessing instructions, and forced recall instructions along with proportion of responses assigned confidence rating of 4 (highest confidence in correct response, confidence ratings of 3 and 4 for critical items only).

	Younger (N = 80)		Older (N = 80)	
	Related	Unrelated	Related	Unrelated
5 Presented Items				
Standard	**.70**	**.49**	**.64**	**.32**
CR 4	.62	.42	.52	.29
Penalty	**.68**	**.45**	**.63**	**.29**
CR 4	.62	.42	.57	.27
No Penalty	**.65**	**.43**	**.66**	**.24**
CR 4	.52	.42	.52	.20
Forced	**.80**	**.48**	**.76**	**.24**
CR 4	.68	.48	.63	.21
5 Critical Items				
Standard	**.16**	**.00**	**.23**	**.00**
CR 3/4	.10	.00	.19	.00
Penalty	**.08**	**.00**	**.19**	**.00**
CR 3/4	.06	.00	.15	.00
No Penalty	**.38**	**.01**	**.31**	**.00**
CR 3/4	.14	.00	.13	.00
Forced	**.46**	**.03**	**.41**	**.01**
CR 3/4	.17	.02	.21	.00
Corrected Recall (Presented - Critical)				
Standard	**.54**	**.49**	**.42**	**.32**

Table 3 Continued

Penalty	**.60**	**.45**	**.46**	**.29**
No Penalty	**.28**	**.42**	**.36**	**.24**
Forced	**.34**	**.46**	**.35**	**.23**

Confidence Ratings First Ten Items (5 Presented)

The proportion of highest confidence ratings (CR 4) assigned by participants to the first five presented items are also presented in Table 3. A separate 2 (Age Group: old vs. young) × 2 (List Type: related vs. unrelated) × 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) mixed factorial ANOVA computed on the mean proportion of highest confidence ratings revealed the same pattern of results as discussed above for the five presented items, demonstrating that high confidence ratings closely matched proportions of correct recall for the five presented items and are not discussed further.

First Ten Items (5 Critical)

Critical intrusions were also examined. Critical intrusions were determined by selecting five of the first ten items for each list and withholding them to assess participants' likelihood of reporting them by guessing during recall. Critical intrusions are reported in Table 3. The proportion of critical intrusions was analyzed using a 2 (Age Group: old vs. younger) × 2 (List Type: related vs. unrelated) × 4 (Test Instructions:

standard vs. penalty vs. no penalty vs. forced) mixed factorial ANOVA with list type as a within-subjects variable. A main effect of list type was found, $F(1, 152) = 567.28$, $MSE = 0.01$, $p < .001$, which demonstrates a greater proportion of related critical items ($M = .27$) were falsely recalled than unrelated critical items ($M = .00$). Recall of critical items from unrelated lists rarely occurred. This is logical because the lack of thematic consistency on unrelated word lists makes producing a critical item very rare because participants could not narrow potential guesses. A main effect of test instructions, $F(3, 152) = 38.87$, $MSE = 0.01$, $p < .001$, indicated that probability of intrusions was contingent on the test instructions. These main effects were also qualified by a significant List Type × Test Instruction interaction $F(3, 152) = 37.32$, $MSE = 0.01$, $p < .001$, and further qualified by a significant List Type × Test Instructions, × Age Group three-way interaction, $F(3, 152) = 2.91$, $MSE = 0.01$, $p = .04$.

Follow up analyses indicated that the three-way interaction was likely attributable to the different pattern of results obtained between younger and older adults across test positions on the related lists (no differences were obtained on the unrelated lists due to floor effects). Younger adults' intrusions appeared to be more flexible dependent upon instruction conditions. Compared to standard instructions ($M = .16$), younger adults reduced the number of critical intrusions when asked to be more conservative in the penalty condition ($M = .08$); however, this difference was only marginally significant $t(38) = 1.76$, $p = .09$. When encouraged to be more liberal in the no penalty and forced conditions (*M*s = .38 and .46, respectively), younger adults increased the number of reported critical intrusions relative to standard instructions (no penalty: $t(38) = 4.01$, *SEM*

= 0.04, $p < .001$; forced: $t(38) = 6.00$, $SEM = 0.04$, $p < .001$). Similarly, younger adults increased intrusions in the liberal conditions relative to the penalty conditions (no penalty: $t(38) = 7.405$, $SEM = 0.02$, $p < .001$; forced: $t(38) = 10.44$, $SEM = 0.02$, $p < .001$. Older adults, on the other hand, were not able to decrease the proportion of critical intrusions under penalty instruction, $t(38) = .99$, $p > .05$. This pattern of results suggests that older adults are not as flexible at reducing the number of critical intrusions as younger adults, and continue to report plausible critical intrusions even under conservative test instructions. Compared to standard instructions, older adults did not increase the proportion of critical intrusions in no penalty test conditions ($M = .31$; $t(38) = 1.92$, $SEM = 0.03$, $p = .06$), although the trend was in this direction. However, older adults did increase their critical intrusions in the forced condition ($M = .41$) relative to the standard, $t(38) = 4.51$, $SEM = 0.03$, $p < .001$. Overall, this pattern of results suggests that older adults were less flexible than younger adults in changing their response criterion in accordance with test instructions.

Evidence that older adults responded more liberally on standard free recall tests comes from age based comparisons across test instructions. Older adults reported a greater number of critical intrusions than younger adults under penalty instructions ($M = .19$ for older adults, $M = .08$ for younger adults, $t(38) = 2.42$, $SEM = 0.02$, $p = .02$), but older and younger adults did not differ in false recall under the most liberal response conditions (forced recall and no penalty, *t*s < 1.0, *p*s > .05). This suggests that older adults were not able to reduce the critical intrusions to comply with test demands under conservative instructions, responding with items that appeared plausible. Younger adults were better able to reduce the number of plausible intrusions they reported and were still able to increase the number of critical intrusions like older adults when the test criterion was relaxed.

Confidence Ratings (Critical Items)

Confidence ratings were analyzed by collapsing across the highest two confidence ratings (3 and 4). Note that CR 3 ratings were included in the current analyses (only CR 4 were analyzed in other sections) because participants frequently utilized confidence ratings of both 3 and 4 when reporting critical intrusions. These data are reported in Table 3. Additionally, confidence ratings were only analyzed for related word lists since critical intrusions for unrelated word lists were absent. A 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) × 2 (Age Group: old vs. young) between-subjects ANOVA was conducted on the proportion of critical intrusions reported with the highest confidence ratings (CR 3 and CR 4). A significant main effect of test instructions, $F(1, 152) = 3.56$, = $MSE = 0.02$, $p < .02$, suggests that high confidence intrusions varied in accordance with test instructions. Specifically, confidence ratings for forced recall ($M = .19$) were significantly greater than penalty instructions ($M = .10$, $t(78) = 3.04$, $SEM = 0.02$, $p = .003$), and no penalty instructions ($M = .13$, $t(78) = 2.02$, $SEM = 0.02$, $p < .05$), but were marginally greater than standard instructions ($M = .14$, $t(78) = 1.79$, $SEM = 0.02$, $p = .08$). There were no differences between other instructions (all ts < 1.24, ps > .05). Critically, a main effect of age group was found, $F(1, 152) = 5.92$, $MSE = 0.02$, $p < .02$, demonstrating that across instructional conditions, older adults ($M = .17$) were more likely to report critical intrusions with higher confidence than younger adults (*M* = .12), a finding consistent with previous research (e.g. Kelley & Sahakyan, 2003). Interestingly, this pattern occurred on free recall tests across instructional manipulations. The interaction between test instructions and age group failed to reach significance, $F(3, 152) = 1.08$, $p > .05$.

A 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) × 2 (Age Group: older vs. younger) between-subjects ANOVA was also completed for the proportion of critical intrusions reported with low confidence (CR 2 and CR 1). A significant main effect of test instructions was found, $F(3, 152) = 32.82$, $MSE = 0.01$, $p < .001$, demonstrating that proportions of low confidence critical intrusions varied based on test instructions. Similar to high confidence critical intrusions, a significant main effect of age group was found, $F(1, 152) = 13.94$, $MSE = 0.01$, $p < .001$, indicating that younger adults were more likely to report low confidence critical intrusions than older adults. These data are complimentary to high confidence critical intrusions demonstrating that younger adults are more flexible with the confidence they have in the correctness of their reported intrusions relative to older adults. These main effects are further qualified by a significant Test Instructions × Age Group interaction, $F(3, 152) = 6.93$, $MSE = 0.01$, $p < .001$. Post hoc analyses were conducted to further explore differences found within this interaction. The only significant difference between older and younger adults found in forced test instructions, with younger adults recalling a greater proportion of low confidence intrusions compared to older adults, $t(38) = 4.29$, $SEM = 0.03$, $p < .001$. All other test instructions failed to reach significance (all *t*s < 1.43, all *p*s > .16). Taken together, evidence of a response bias between older and younger adults can be found with the confidence levels of reported critical intrusions. Critically, younger adults were more likely to report low confidence critical intrusions than older adults, but only when required to guess under forced recall instructions. Since older adults reported fewer low confidence intrusions, this suggests that they report higher confidence intrusions by default. Thus, older adults appear to be overconfident in the accuracy of their intrusions.

First Ten Items (Corrected Recall)

Corrected recall was calculated for the first ten items by subtracting the proportion of recall of the five critical items from the proportion of recall from the five presented items. The use of corrected recall allows for the assessment of memory performance without the influence of erroneous recall (corrected recall is presented in Table 3). Corrected recall was subjected to a 2 (Age Group: old vs. younger) × 2 (List Type: related vs. unrelated) × 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) mixed factorial ANOVA with list type as a within-subjects variable. A main effect of list type was found, $F(1, 152) = 12.75$, $MSE = 0.02$, $p < .001$, indicating that related list corrected recall ($M = .42$) was greater overall than unrelated corrected recall ($M = .36$). Additionally, a significant main effect of test instructions, $F(3, 152) = 9.19$, $MSE = 0.03$, $p < .001$, demonstrated differences in corrected recall depended on instructions. These main effects were further qualified by a significant List Type × Test Instructions interaction, $F(3, 152) = 8.87$, $MSE = 0.02$, $p < .01$, indicating greater differences between test instructions in related lists as opposed to unrelated lists. A significant main effect of age group, $F(1, 152) = 33.96$, $MSE = 0.03$, $p < .001$, indicates that younger adults had a higher level of corrected recall than older adults. This main effect was qualified by a significant Age Group × List Type interaction, $F(1, 152) = 35.99$, $MSE = 0.02$, $p < .001$, and further qualified by a significant List Type × Test Instructions × Age Group three-way interaction, $F(3, 152) = 5.46$, $MSE = 0.02$, $p = .001$.

Follow up comparisons indicated that test instructions did not influence corrected recall for unrelated word lists (all *t*s < 1.68, *p*s > .09). Therefore, post hoc analyses were carried out to further explore differences but only on related word lists. Namely, younger adults' corrected recall under standard instructions ($M = .54$) was equal to penalty instructions ($M = .60$, $t(38) = 1.04$, $SEM = 0.04$, *ns*.), but greater than no penalty instructions ($M = .28$, t(38) = 4.06, $SEM = 0.04$, $p < .001$) and forced instructions ($M = .34$, $t(38) = 3.32$, $SEM = 0.04$, $p = .002$). Similar to younger adults, the proportion of corrected recall for older adults in standard instructions ($M = .42$) was equal to penalty instructions ($M = .46$, $t(38) = .73$, $SEM = 0.03$, *ns*.). Critically, older adults' corrected recall in standard conditions was also equal to no penalty instructions ($M = .36$, $t(38) = 1.34$, $SEM = 0.03$, $p > .05$), as well as forced recall ($M = .35$, $t(38) = 1.41$, $SEM = 0.04$, $p > .05$). This pattern demonstrates important response differences between older and younger adults. For younger adults, their corrected recall closely mimicked instructional conditions, with higher levels of corrected recall in conservative conditions where they were encouraged to withhold intrusions and lower corrected recall in conditions in which they were encouraged to guess. In contrast, older adults' corrected recall under standard conditions was equivalent to their corrected recall under liberal response conditions. In essence, the natural memory responding for older adults as demonstrated under standard free recall instructions is equivalent to liberal recall instructions because older adults were likely to report more plausible critical intrusions without instructions to do so.

Total Number of Intrusions

Converging evidence for our conclusions that older adults are reporting additional items comes from the total number of intrusions which include the critical intrusions discussed above and any additional extralist intrusions. The total number of intrusions reported per list are presented in Table 4. The average total number of intrusions per list was submitted to a 2 (Age Group: young vs. old) × 2 (List Type: related vs. unrelated) × 4 (Test Instructions: standard vs. penalty vs. no penalty vs. forced) mixed factorial ANOVA with list type as a within-subjects variable. Significant main effects were found for test instructions, $F(3, 152) = 22.17$, $MSE = 1755.30$, $p < .001$, and list type, $F(1, 152) = 194.97$, $MSE = 414.62$, $p < .001$, demonstrating that participants varied the number of reported intrusions depending on the instructions given at recall and the total number of reported intrusions were fewer when participants recalled related word lists ($M = 3.59$) compared to unrelated word lists ($M = 5.86$). The ANOVA also revealed a significant List Type × Test Instruction interaction, $F(3, 152) = 107.70$, $MSE = 2.13$, $p < .001$, which was further qualified by a significant three-way interaction, $F(3, 152) = 3.46$, $MSE = 2.13$, $p = .02$.

Post hoc analyses were carried out on the total number of intrusions reported to examine the three-way interaction. Specifically, when recalling from related lists, younger adults reported significantly fewer intrusions under standard (*M* = 1.24) and penalty instructions (*M* = .65, *t*(38) = 2.28, *SEM* = .31, *p* = .03; *t*(38) = 2.83 *SEM* = .14, *p* < .01, respectively) compared to older adults (for standard: *M* = 2.30, for penalty: *M* = 1.51). However, younger and older adults did not differ in the number of extralist intrusions reported under no penalty instructions (for younger adults, *M* = 3.95, for older adults *M* = 3.46, *t*(38) = .76, *SEM* = 0.40, *p* = .45), or forced recall instructions (for younger adults, *M* = 7.70, for older adults *M* = 7.88, *t*(38) = .32, *SEM* = 0.25, *p* = .75). In other words, when guessing was encouraged or required, older and younger adults produced equivalent numbers of extralist intrusions. When guessing was discouraged, younger adults were better able to reduce the number of extralist intrusions than were older adults, suggesting that older adults are approaching standard and penalty conditions with a greater propensity toward guessing than are younger adults.

Additionally, for the unrelated lists under forced recall instructions, older adults (*M* = 16.11) reported more intrusions than younger adults (*M* = 13.93, *t*(38) = 3.50, *SEM* = 0.44, *p* = .001). This was the only significant difference between age groups and test instructions (all test instruction differences *t*s < 1, *p*s > .05). This pattern of results is not surprising since older adults were not able to recall as many correct items after studying an unrelated word list and they were then required to report additional intrusions to meet task demands.

These data suggest an interesting pattern with regards to a differing response bias between older and younger adults. Specifically, on recall tests that encouraged guessing, younger and older adults did not differ in the number of extralist intrusions reported.

However, on standard recall tests, older adults responded more liberally as evidenced by a greater willingness to report incorrect memory items. Additionally, this tendency to report incorrect items persists into conditions where a strong penalty is associated with an incorrect response. Oppositely, younger adults were better able to reduce the total number of reported intrusions under standard and penalty test instructions.

Table 4. Mean total number of intrusions incorrectly recalled per list by younger and older adults on recall tests with standard instructions, penalty for guessing instructions, no penalty for guessing instructions, and forced recall instructions (in bold) along with proportion of responses assigned confidence rating of 4 (highest confidence in correct response).

	Younger (N = 80)		Older (N = 80)	
	Related	Unrelated	Related	Unrelated
Standard Total	**1.24**	**1.51**	**2.30**	**2.28**
CR 4	.20	.19	.44	.31
Penalty Total	**.65**	**1.14**	**1.51**	**1.39**
CR 4	.20	.33	.50	.36
No Penalty	**3.95**	**5.51**	**3.46**	**4.99**
CR 4	.50	.30	.48	.36
Forced	**7.70**	**13.93**	**7.88**	**16.11**
CR 4	.58	.35	.84	.50

Correlations with Memory Anxiety Measures

Individual performance differences on memory tests were correlated with the memory anxiety questionnaire (Davidson et al., 1991) to further assess the relationship between memory anxiety and response biases. Specifically, the relationship between the average score on the memory anxiety questionnaire and correct recall and the average number of intrusions reported per list were examined. Memory anxiety was marginally correlated with correct recall performance on related word lists, $r = -.15$, p(two-tailed) = .06, but was significantly correlated with correct recall on unrelated word lists, $r = -.23$, $p < .01$, demonstrating a negative relationship between memory anxiety and correct recall performance across older and younger adults. For total number of intrusions recalled by both age groups, anxiety was marginally related to the intrusions produced when recalling from a related word lists, $r = .15$, $p = .06$, but not related to an unrelated word list, $r = .09$, $p > .05$. Therefore, higher scores on the anxiety measure were associated with lower levels of correct recall for unrelated list and higher levels of intrusions for related lists.

Further, correlational analyses were also conducted on the final question on the memory anxiety questionnaire which assessed an individual's subjective rating of the importance of reporting additional information during recall rather than withholding information. This question is of particular importance to the current study since a response bias is indicative of reporting differences found in intrusion rates. For proportions of correct recall, correlational analyses did not reveal significant relationships with responses for question 12 on the anxiety questionnaire ($r = -.04$ for related word list correct recall, $r = -.10$ for unrelated word list correct recall, both ps > .23). However, a significant positive correlation was found with total intrusions reported on related word lists ($r = .20$, $p < .02$), and a marginal correlation was found with total intrusions on unrelated word lists ($r = .14$, $p = .10$), suggesting that participants' beliefs in reporting additional items may have resulted in additional intrusions.

To delineate response differences between older and younger adults, correlational analyses for question 12 on the memory anxiety questionnaire were completed for both age groups separately. Analyses demonstrated no significant relationships between anxiety and correct recall for both related and unrelated word lists with older adults (for related, $r = -.11$, $p > .33$, for unrelated, $r = -.01$, $p > .90$), and younger adults (for related, $r = .11$, $p > .35$, for unrelated, $r = -.13$, $p > .29$). Surprisingly, correlational analyses revealed a significant positive relationship with younger adults for the total number of intrusions reported per list (for related, $r = .38$, $p < .001$, for unrelated, $r = .34$, $p < .01$), but not for older adults (for related, $r = .03$, $p > .80$, for unrelated, $r = -.03$, $p >.80$). This pattern operates counter to reporting patterns found above, demonstrating that older adults were more likely to report intrusions than younger adults. While correlational analyses should be interpreted with caution, it is important to acknowledge that anxiety measures were completed at the end of the study and may be a result of retrospective processes as participants reflected on their performance during the experimental session, rather than typical memory performance. Therefore, older adults may not be able to retrieve their performance during the experimental session which would explain the lacking relationship involving older adults. While the present study is unable to distinctly determine causes behind these patterns, one potential area of interest for future research may be to address the discrepancy between older and younger adults behavioral reporting and their responding in regards to intrusions as assessed by the memory anxiety questionnaire.

DISCUSSION

The present experiment provides the first examination of possible age-related response biases on free and forced recall tests. Specifically, older and younger adults may differentially report items for which there is a large or small amount of evidence that they were previously presented during study. Based on this logic, a response bias is operationalized as the number of intrusions that individuals are willing to report. Comparing age group differences across categorized and unrelated word lists differing in contextual support, as well as test instructions designed to encourage or discourage intrusion reporting, response tendencies could be assessed resulting in novel findings regarding the willingness to report memory items and confidence in the accuracy of such items.

Correct Recall

Proportions of correct recall indicated that participants were able to adjust their recall levels in accordance with test instructions, but only when recalling from a categorized word list. This pattern of data demonstrates that participants are able to strategically regulate memory performance by increasing correct recall by guessing when encouraged to do so, and also reducing the proportion of correct recall under penalty test instructions. These data are consistent with the literature that participants are able to monitor their memory responses for accuracy (Koriat & Goldsmith, 1996). Interestingly, the current study found that older and younger adults adjust correct recall to the same magnitude.

On unrelated word lists, younger adults outperformed older adults, nearly doubling their proportion of correct recall. This pattern is not surprising and has been well documented as older adults appear to be more sensitive to contextual support for memory performance (cf. Balota et al., 2000). Surprisingly, younger adults also outperformed older adults on categorized word lists which contained increased contextual support. Research has demonstrated that when study and test materials are used that increase the availability of contextual information (e.g., categorized items, cued recall tests, recognition tests), age differences in memory performance are often greatly reduced or eliminated (Craik & McDowd, 1987; Rabinowitz et al., 1982). Although categorized materials reduced age differences, our data demonstrate that older adults reliably underperform younger adults on categorized word lists, but this difference is relatively small (about 4.5%). Similarly, these data appear consistent with other studies using the Koriat and Goldsmith (1996) metamemory paradigm, with older adults recalling fewer correct items in both forced and free-report cued recall (Kelley & Sahakyan, 2003; Pansky et al., 2009).

Additionally, confidence was assessed for reported responses. Across test instructions and age groups, high confidence ratings closely mirrored the proportion of correct responses. This suggests that participants were aware that the items reported were correct. Confidence ratings have rarely been assessed on free recall tasks, as it has been assumed that responses are reported with the highest confidence that the item was previously studied. Based on the correspondence between high confidence ratings and the proportion of correct recall, this assumption appears to be warranted. In regards to differential responding between older and younger adults, both groups rated correct items with similar confidence. This finding is consistent with Dehon and Bredart (2004), who showed that younger and older adults demonstrated equivalent ratings on studied items correctly recalled from the Deese-Roediger-McDermott (DRM) false memory paradigm (Deese, 1959; Roediger & McDermott, 1995). However, as mentioned previously, a response bias is assessed through the willingness to report from memory yielding differences in intrusions likely varying in subjective assessments.

Intrusions

An individuals' willingness to report memory items is measureable primarily in the intrusions provided on a free recall task. Intrusions are particularly sensitive to a response bias because if individuals respond differently, this will manifest in errors due to guessing. Data from the overall extralist intrusions indicates that participants were successfully able to increase or decrease the number of extralist intrusions based on task demands. The question of interest for the present experiment was how age groups are willing to report intrusions and if they differed based on task instructions. When presented with related word lists, both younger and older adults adjusted the number of intrusions, but younger adults were more flexible in the number of incorrect responses provided. Older adults appeared less sensitive to recall instructions and responded with a relatively similar number of intrusions across test instructions. A similar pattern occurred with unrelated lists.

Critical differences in response biases were found in the number of intrusions reported under standard instructions. Standard instructions did not encourage participants to guess or withhold incorrect responding and therefore was an assessment of an individual's natural memory responding without task constraints. Under standard instructions, older adults reported a greater number of intrusions relative to younger adults, demonstrating that older adults are biased to respond liberally when reporting from memory on a free recall task. This pattern of results demonstrates that older adults report memory items with a liberal bias and were naturally more likely to guess than younger adults. Younger adults, on the other hand, were relatively conservative in their responding, minimizing their intrusions even if they were not directly instructed to do so.

A response bias was further corroborated with performance data on the guess-check procedure on categorized word lists. The guess-check procedure allowed for the assessment of critical guessable intrusions that were most plausible of a given category, but were not presented to participants. If individuals were to guess, they would likely constrain their guesses to the presented category and produce the most plausible nonpresented critical items. Therefore, this procedure can distinguish items produced as guesses from items retrieved from memory.

Guess-Check Procedure

The present study finds support for differential response biases between older and younger adults using the guess-check procedure. Similar to patterns found in overall extralist intrusions, younger adults were more flexible in reporting the plausible critical intrusions across test instructions compared to older adults on categorized word lists. When tested in penalty conditions, older adults were less likely to reduce the number of reported critical items than younger adults. This pattern is consistent with other research demonstrating that when given an option to reduce intrusions; older adults do not minimize their erroneous responses compared to younger adults (Jacoby et al., 2005). This pattern is particularly alarming because it demonstrates that the liberal response bias found in older adults is so pervasive that, even when instructed, older adults are not able to reduce their intrusions. This tendency is the strongest piece of evidence that older adults are liberal when reporting responses from memory. Conversely, younger adults are able to reduce the number of intrusions based on instruction and are capable of responding more conservatively. However, when younger and older adults are instructed to produce intrusions, they do so equally.

Confidence ratings for critical intrusions were also examined to determine if a response bias was also demonstrated through differing phenomenology. Critically, older adults were more likely than younger adults to report critical items with high confidence, consistent with both metamemory and false memory literature (Kelley & Sahakyan, 2003; Dehon & Bredart, 2004). This suggests that older adults may not be aware of their response bias and are overconfident in the items they report from memory. This type of responding is particularly interesting because even in liberal test instructions encouraging participants to report intrusions, older adults report intrusions with high confidence. Across test instructions, younger adults reported critical intrusions with less confidence. Taken together, younger adults are biased to respond relatively conservatively when reporting memory items, minimizing intrusions, but when intrusions are recalled under forced choice conditions, they have little confidence in their accuracy. Older adults demonstrate a relatively liberal response bias, recalling a greater number of intrusions, and having greater confidence that their intrusions were previously studied, even under forced report conditions.

Limitations

One potential explanation for these results may be that younger and older adults were not equated at initial encoding, resulting in differences in responding. Prior research has demonstrated that providing older adults with more time to encode study items does increase their recall performance, minimizing differences between age groups (Salthouse, 1994). Although a response bias could be due to either encoding or retrieval processes, there is some evidence demonstrating that the same liberal response bias found in older adults still occurs when younger adults are equated to older adults in quantitative performance by dividing younger adults attention during encoding (Pansky et al., 2009). In the present study, the quantitative performance between younger and older adults was more similar on categorized word lists relative to unrelated word lists suggesting that younger and older adults had similar levels of original learning, and a liberal response bias was still measured. If different levels of encoding could influence the criteria one uses to report from memory, future research endeavors could manipulate the encoding ability of not only younger adults, but also older adults by increasing the encoding time given to older adults. Differential memory representations could influence not only how individuals report items from memory but also the confidence in which memories are reported.

Future Implications

The existence of a response bias between older and younger adults has ramifications for future research endeavors. Memory experiments examining age differences may need to equate for biases to equate for responding, but only when stimuli are thematically consistent, similar to our related word lists. For example, researchers may need to inform older adults to withhold guesses to equalize responding with younger adults. Oppositely, informing younger adults to be liberal in their responding may match the number of intrusions naturally produced by older adults. However, it may be more difficult to equate to older adults as younger adults are more flexible with their intrusions and therefore, may actually produce more intrusions than older adults report naturally.

Conclusion

In summary, the goal of the present research was to explore possible age differences between younger and older adults in responding on free and forced recall tasks that may be involved in performance differences when recalling from memory. First, both older and younger adults can increase or decrease the number of correct items recalled depending on testing instructions, but only when individuals studied a categorized word list. This pattern demonstrates that individuals are able to deliberately control their own individual levels of recall, but only when they have contextual support that allows them to differentially increase recall. The increase in correct recall comes at the expense of intrusions. Second, a response bias as indicated by the number of intrusions reported by older and younger adults, demonstrated age differences in responding. Older adults were more likely to report intrusions even under conditions in which guesses are penalized. The increased levels of intrusions appeared to be particularly high for older adults when they were plausible given a categorized study list. Finally, older adults reported intrusions with greater confidence than younger adults. This demonstrates that older adults may not necessarily be aware that an item they reported was incorrect and perhaps that older adults are overconfident in their own memory abilities. Thus the liberal response bias demonstrated by older adults is marked by the increased number of intrusions provided and the increased level of confidence that the reported items were studied.

REFERENCES CITED:

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science, 220,* 25-30. doi: 10.1126/science.6828877

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97-123. doi: 10.1037/h0033773

Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: I. Evidence from divided attention costs. *Psychology and Aging, 13,* 405-423. doi: 10.1037/0882-7974.13.3.405

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory in healthy older adults. In F. I. M. Craik & E. Tulving (Eds.), *The Oxford handbook of memory* (pp. 411-425). New York: Oxford University Press.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445-459.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80,* 1-45. doi: 10.1037/h0027577

Cofer, C. N. (1967). Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), Concepts and the structure of memory. New York: Wiley.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497-505.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 474-479. doi: 10.1037/0278-739313.3.474

Davidson, H. A., Dixon, R. A., & Hultsch, D. F. (1991). Memory anxiety and memory performance in adulthood. *Applied Cognitive Psychology, 5,* 423-433. doi: 10.1002/acp.2350050504

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17-22. doi: 10.1037/h0046671

Dehon, H., & Bredart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology and Aging, 19,* 191-197. doi: 10.1037/0882-7974.19.1.191

Erdelyi, M. H., Finks, J. & Feigin-Pfau, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General, 118,* 245-254. doi: 10.1037/0096-3445.118.3.245

Folstein, M.F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical methodfor grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189-198. doi: 10.1016/0022-3956(75)90026-6

Henkel, L. A. (2007). The benefits and costs of repeated memory tests for young and older adults. *Psychology and Aging, 22,* 580-595. doi: 10.1037/0882-7974.22.3.580

Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General, 134,* 131-148. doi: 10.1037/0096-3445.134.2.131

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48,* 704-721. doi: 10.1016/S0749-596X(02)00504-1

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517. doi: 10.1037/0033- 295X.103.3.490

Luszcz, M. A., Roberts, T. H., & Mattiske, J. (1990). Use of relational and item-specific information in remembering by older and younger adults. *Psychology and Aging, 5,* 242- 249. doi: 10.1037/0882-7974.5.2.242

McKelvie, S. J. (2001). Effects of free and forced retrieval instructions on false recall and recognition. *Journal of General Psychology, 128,* 261-278.

Meade, M. L., & Roediger, H. L., III. (2006). The effect of forced recall on illusory recollection in younger and older adults. *The American Journal of Psychology, 119,* 433-462.

Meade, M. L., & Roediger, H. L., III. (2009). Age differences in collaborative memory: The role of retrieval manipulations. *Memory & Cognition, 37,* 962-975. doi: 10.3758/MC.37.7.962

Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N., (1989). Effects of age and divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 6,* 1185-1191. doi: 10.1037/0278-7393.15.6.1185

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., & Gaines, C. L. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11,* 621-637. doi: 10.1037/0882-7974.11.4.621

Pansky, A., Goldsmith, M., Koriat, A., Pearlman-Avnion, S. (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology, 21*, 303-329. doi: 10.1080/09541440802281183

Rabinowitz, J. C., Craik, F. I. M., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology, 39,* 325 344. doi: 10.1037/h0080643

Ritter, W. & Buschke, H. (1974). Free, forced and restricted recall in verbal learning. *Journal of Experimental Psychology, 6,* 1204-1207. doi: 10.1037/h0037411

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803-814.doi: 10.1037/0278-7393.21.4.803

Roediger, H. L., III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesias: A puzzle for generate/recognize theories. *Memory & Cognition, 13,* 1-7.

Roediger, H. L., III, Srinivas, K., & Waddill, P. (1989). How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. *Journal of Experimental Psychology: General, 118,* 255-257. doi: 10.1037/0096-3445.118.3.255

Roediger, H. L., III, Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing, and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 97-134). San Diego, CA: Academic Press.

Ronnlund, M., Nyberg, L., Backman, L., & Nilsson, L. (2003). Recall of subject performed tasks, verbal tasks, and cognitive activities across the adult life span.

Aging, Neuropsychology, and Cognition, 10, 182-201. doi: 10.1076/anec.10.3.182.16449

Salthouse, T. A. (1994). Aging associations: Influence of speed on adult age differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1486-1503. doi: 10.1037/0278-7393.20.6.1486

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9,* 371-377.

Veiel, L. L., & Stroandt, M. (2003). Processing costs of semantic and episodic retrieval in younger and older adults. *Aging, Neuropsychology, and Cognition, 10,* 61-73. doi: 10.1076/anec.10.1.61.13458

APENDICES:

APPENDIX A:

CATEGORIZED AND UNRELATED STUDY LISTS

Categorized Study Lists

A	B	C	D
1. dog*	1. legs*	1. robin*	1. hammer*
2. cat^	2. arms^	2. sparrow^	2. saw^
3. horse^	3. head^	3. cardinal^	3. nails^
4. cow*	4. eye*	4. eagle*	4. screwdriver*
5. lion*	5. foot*	5. crow*	5. level*
6. tiger^	6. nose^	6. bluebird^	6. plane^
7. elephant^	7. finger^	7. canary^	7. chisel^
8. pig*	8. ear*	8. parakeet*	8. ruler*
9. bear*	9. hand*	9. hawk*	9. wrench*
10. mouse^	10. toe^	10. blackbird^	10. pliers^
11. rat	11. mouth	11. wren	11. drill
12. deer	12. stomach	12. oriole	12.screws
13. sheep	13. hair	13. parrot	13. pencil
14. giraffe	14. neck	14. pigeon	14. square
15. goat	15. heart	15. hummingbird	15. sawhorse
16. zebra	16. knee	16. starling	16. file
17. squirrel	17. chest	17. woodpecker	17. lathe
18. wolf	18. liver	18. vulture	18. sandpaper
19. donkey	19. brain	19. swallow	19. awl
20. rabbit	20. lungs	20. chicken	20. crowbar
21. leopard	21. tooth	21. dove	21. sander
22. mule	22. elbow	22. duck	22. wood
23. fox	23. shoulders	23. owl	23. plumb
24. bull	24. face	24. redbird	24. wedge
25. buffalo	25. tongue	25. thrush	25. axe

* and ^ denotes 5 items presented or withheld (critical items) that were used in the guess-check procedure.

Unrelated Study Lists

A	B	C	D
1. key*	1. wood*	1. trash*	1. bullet*
2. sea^	2. snow^	2. skillet^	2. air^
3. wheel^	3. land^	3. tortoise^	3. knife^
4. pie*	4. bar*	4. rifle*	4. rectangular*
5. whip*	5. wire*	5. hood*	5. field*
6. ranch^	6. sand^	6. lollipop^	6. river^
7. triangle^	7. engine^	7. oyster^	7. palace^
8. zoo*	8. bag*	8. skydivers*	8. crypt*
9. lake*	9. book*	9. kite*	9. piston*
10. stick^	10. pie^	10. waterfall^	10. pickle^
11. hat	11. train	11.beet	11. shore
12. leaf	12. trumpet	12. lichen	12. walnut
13. piano	13. park	13. pillow	13. chapel
14. thistle	14. salt	14. maroon	14. soccer
15. cone	15. radio	15. thermometer	15. sunrises
16. yacht	16. shoe	16. tricycle	16. mail
17. textbook	17. crown	17. lumberjack	17. kayak
18. deck	18. metal	18. sunroof	18. toothpick
19. shrimp	19. shell	19. gallery	19. oaf
20. ribbon	20. towel	20. mailbox	20. bandage
21. noodles	21. lemon	21. foam	21. racket
22. vine	22. fence	22. horn	22. bomb
23. jet	23. hurricane	23. fur	23. easel
24. dirt	24. male	24. skewers	24. pasta
25. stadium	25. garden	25. argon	25. hut

* and ^ denotes 5 items presented or withheld (critical items) that were used in the guess-check procedure.

APPENDIX B:

MEMORY ANXIETY QUESTIONNAIRE

Appendix B

Memory Anxiety Questionnaire (Davidson, Dixon, & Hultsch, 1991).
Respond:

5=Strongly Agree
4=Agree
3=Neutral
2=Disagree
1=Strongly Disagree

1. I would feel on edge right now if I had to take a memory test.

2. When someone I don't know very well asks me to remember something I get nervous.

3. I am usually uneasy when I attempt a problem that requires me to use my memory.

4. I get tense and anxious when I feel my memory is not as good as other people's.

5. I get anxious when I am asked to remember.

6. I do not get flustered when I am put on the spot to remember new things.

7. I feel jittery if I have to introduce someone I just met.

8. I get anxious when I have to do something that I haven't done in a long time.

9. If I am put on the spot to remember names, I know I will have difficulty doing it.

10. I would feel very anxious if I visited a new place and had to remember how to find my way back.

11. I get upset when I cannot remember something.

12. When taking a memory test, I feel it is a more serious memory error to leave something out than it is to write down something extra.