Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Memory and Language 52 (2005) 578–594

Journal of Memory and Language

www.elsevier.com/locate/jml

Executive processes, memory accuracy, and memory monitoring: An aging and individual difference analysis ☆

Matthew G. Rhodes *, Colleen M. Kelley

Department of Psychology, Florida State University, Tallahassee, FL 32306-1270, USA

Received 31 August 2004; revision received 21 January 2005
Available online 3 March 2005

Abstract

The current study examined the neuropsychological correlates of memory accuracy in older and younger adults. Participants were tested in a memory monitoring paradigm developed by Koriat and Goldsmith (1996), which permits separate assessments of the accuracy of responses generated during retrieval and the accuracy of monitoring those responses. Participants were also administered a battery of tests designed to measure executive functioning and speed of processing. Results indicated that both age and executive measures were predictive of accuracy, while speed of processing measures accounted for little of the variability in accuracy. Path analyses demonstrated that a substantial portion of the effect of executive function measures on memory accuracy in free report was mediated by the quantity of correct responses available in forced report, which in turn was partially mediated by monitoring accuracy. These data suggest that individual differences in executive function are important in memory accuracy.

Keywords: Memory; Metamemory; Memory monitoring; Aging; Executive function

---

Introduction

The quality of one's memory has traditionally been characterized in terms of the quantity of ideas or the number of aspects of events that are recalled. However, memory accuracy has been the subject of growing interest (see Koriat, Goldsmith, & Pansky, 2000; Roediger, 1996 for reviews) with a particular focus on errors of commission. For example, memory accuracy can be quite low after the introduction of misleading post-event information (e.g., Loftus, Miller, & Burns, 1978), after studying texts that introduce strong inferences (e.g., Owens, Bower, & Black, 1979), and in paradigms such as the Deese–Roediger–McDermott (Deese, 1959; Roediger & McDermott, 1995), or DRM paradigm, in which studying lists of associates of a central, nonpresented item induces high levels of false recall. In

---

☆ Parts of this research were included in a Dissertation submitted to The Florida State University by Matthew G. Rhodes. We thank the other members of the dissertation committee, Michelle Bourgeios, Neil Charness, Katinka Dijkstra, and Ashby Plant for their helpful comments throughout the course of the research. We also thank John Dunlosky and Morris Goldsmith for their comments and suggestions on an earlier version of the manuscript. We are indebted to Matthew Campioni, Justin Crowe, Mary Currie, Kristy McDaniel, Kathleen Muller, Lisa Spaulding, and Cicely Procipio for their assistance in testing subjects.

* Corresponding author. Present address: Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130, USA.

E-mail address: mrhodes@artsci.wustl.edu (M.G. Rhodes).

0749-596X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2005.01.014

the current study, we use an individual differences analysis across younger and older adults to assess the role of executive functions in memory accuracy and suggest that they are related to both the quantity of correct information available to the rememberer and the monitoring of this information.

Executive functions have been specified in a number of ways and may include the monitoring and control of behavior, suppression of irrelevant information, reasoning, updating information in working memory, inhibition of prepotent responses, planning, shifting, and control of attention, among others (e.g., Baddeley, 1996; Fisk & Sharp, 2004; Kane & Engle, 2002; Miyake et al., 2000; Shimamura, 2000a, 2000b; Waltz et al., 1999). Several researchers have suggested that a decline in executive functions is a primary factor in the cognitive deficits present in aging populations (e.g., Moscovitch & Winocur, 1995; West, 1996; Whelihan & Lesher, 1985). Executive functions are presumably localized in the prefrontal cortex (e.g., Duncan, 1995; Kane & Engle, 2002; Shallice & Burgess, 1991; Waltz et al., 1999), which deteriorates more rapidly with age than other cortical regions (see Raz, 2000, for a review). There is extensive behavioral evidence indicating that older adults perform more poorly than young adults on a number of tasks thought to tap executive functioning that are likewise sensitive to frontal lobe lesions. For example, older adults make more perseverative errors on the Wisconsin Card Sorting Test (WCST; see Rhodes, 2004, for a review), show greater interference on incongruent trials of the Stroop task (Houx, Jolles, & Vreeling, 1993), and produce fewer words on tests of verbal fluency (e.g., Howard, 1980).

The motivation for the current study is a set of findings indicating that variations in performance on neuropsychological tests of executive function predict the memory accuracy of older adults (e.g., Butler, McDaniel, Dornburg, Roediger, & Price, in press), who generally exhibit more false memories than younger adults (e.g., Bartlett, Halpern, & Dowling, 1995; Jacoby, 1999a; Kelley & Sahakyan, 2003; Norman & Schacter, 1997). For example, Butler et al. (in press) demonstrated that older adults scoring poorly on tests of executive function were highly likely to make errors of commission during recall in the DRM paradigm. In contrast, older adults with relatively high scores on measures of executive function exhibited levels of accuracy comparable to young adults. This is consistent with data from other memory tasks showing significantly better memory performance by older adults scoring high on measures of executive functioning in comparison to groups of older adults with relatively poor performance on such measures (e.g., Davidson & Glisky, 2002; Glisky, Polster, & Routhieaux, 1995; Glisky, Rubin, & Davidson, 2001; McDaniel, Glisky, Rubin, Guynn, & Routhieaux, 1999; see also Crawford, Bryan, Luszcz, Obonsawin, & Stewart, 2000).

While executive processes are predictive of memory accuracy, the specific nature of this relationship is not entirely clear. We consider three possibilities. First, executive functions may be crucial to the encoding and retrieval of accurate information. This may include binding information during encoding (cf., Chalfonte & Johnson, 1996; Glisky et al., 2001; Henkel, Johnson, & De Leonardis, 1998) or maintaining an appropriate retrieval set that allows one to generate cues that are likely to lead to the retrieval of accurate information (Burgess & Shallice, 1996; Moscovitch & Melo, 1997; Norman & Schacter, 1996). For example, Norman and Schacter (1996) have proposed that high levels of false memories observed in frontal patients occur either because they are totally unable to recapitulate the study context or because they produce "unfocused retrieval descriptions" that provide only a vague representation of the study context. False memories in older adults may also be a consequence of relying on vague or unfocused representations of study episodes that lack specific details indicative of prior occurrence (e.g., Jacoby, 1999a, 1999b; Jennings & Jacoby, 1997; Kelley & Sahakyan, 2003; see also Parkin & Walter, 1992). Thus, executive processes may in part mediate encoding and/or retrieval processes that are important for memory accuracy.

Second, executive functions may be important for monitoring candidate responses for accuracy (Shimamura, 2000a, 2000b). Monitoring and controlling the contents of cognitive processes likely comprises an important executive function (cf. Fernandez-Duque, Baird, & Posner, 2000) that has implications for the attainment of accuracy (e.g., Shimamura, 2000a, 2000b). Evidence from frontal patients indicates that they have considerable difficulty with complex metacognitive tasks such as memory monitoring (e.g., Janowsky, Shimamura, & Squire, 1989; Vilkki, Servo, & Surma-aho, 1998; see Shimamura, 1996 for a review). In addition, there is some evidence that measures of executive function are predictive of certain types of memory monitoring judgments. For example, Souchay, Isingrini, and Espagnet (2000) reported that the accuracy of feeling-of-knowing judgments was positively correlated with performance on the WCST and verbal fluency tasks (see also Souchay, Isingrini, Clarys, Taconnat, & Eustache, 2004). Given this, any relationship between accuracy and measures of executive function may in part reflect memory monitoring.

Third, these factors may not be independent. That is, the relationship between executive functions and monitoring may be partially mediated by the quality or quantity of candidate responses available in memory. Kelley and Sahakyan (2003) suggest that monitoring depends largely on the quality of information available to the monitoring process, such that monitoring will be more difficult if participants retrieve vague or undifferentiated information. For example, they demonstrated that the poorer monitoring performance of older adults in their study of memory for paired associates was also apparent for young adults whose encoding was disrupted by a divided attention task (cf. Jacoby, 1999a). Kelley and Sahakyan argued that dividing attention at encoding reduced the probability that participants could recollect prior details of studied items, and thus reduced the quality of evidence that candidate responses were indeed memories. Regression analyses of memory accuracy indicated that both the quantity of memory available and a measure of monitoring accounted for significant, unique variability in memory accuracy, but there was substantial shared variance as well.

We propose a theoretical framework whereby individual differences in executive functions produce variations in memory accuracy that are at least partially mediated by effects on the candidate responses people generate, either because of poor encoding or unfocused retrieval processes. The effects of executive functions on the candidate responses generated may, in turn, affect the ability to monitor the correctness of candidate responses. In addition, individual differences in executive functions may also partially mediate the ability to monitor candidate responses, independent of effects on the candidate responses that are retrieved.

The current study

The current study used the memory monitoring framework developed by Koriat and Goldsmith (1994, 1996) to examine the relationship between measures of executive function and memory accuracy. Koriat and Goldsmith's framework is ideal for investigating such issues as it permits a separate assessment of the quantity of retrieved responses and the effectiveness of memory monitoring in the attainment of memory accuracy. In particular, Koriat and Goldsmith suggest that memory performance is not simply a function of retrieving a latent memory trace, but, in addition, depends on the degree to which the rememberer can monitor the correctness of responses, uses this information to select a correct response, and sets an appropriate criterion based on incentives for accuracy. An important component of this perspective is that it distinguishes between quantity and accuracy-based measures of performance. Specifically, quantity is measured as the number of correct responses elicited during an initial, forced recall stage. However, monitoring and control processes may then operate during a subsequent, free recall stage, permitting participants to control the accuracy of their output. Accuracy can thus be measured as the proportion of correct responses volunteered out of the number of responses offered. In particular, Koriat and Goldsmith's framework suggests that participants monitor the correctness of a candidate response by assigning it a probability of being correct (P_a). They then control their responding by setting a response criterion, P_{rc}. If P_a exceeds or is equal to P_{rc} an answer is volunteered; otherwise, the response is withheld. The criterion (P_{rc}) can be adjusted based on payoffs specifying the relative costs and benefits for correct and incorrect responses. Consequently, performance at free report is dependent on three factors. The first is *monitoring effectiveness*, which is the extent to which assessed probabilities of being correct (i.e., P_a) successfully distinguish between correct and incorrect responses. A second factor is *control sensitivity*, which captures the degree to which participants base their decision to respond on the assessed probability of being correct. Finally, participants may adjust their response criterion depending on incentives for accuracy (*response criterion setting*).

Kelley and Sahakyan (2003) used Koriat and Goldsmith's (1994, 1996) framework to investigate memory accuracy and monitoring performance in older and younger adults in paired associate, cued recall. Results showed that older adults were significantly less likely to produce a correct response during the forced report stage (i.e., quantity) and, as well, were less accurate than young adults for those responses that they chose to volunteer during the free report phase (i.e., accuracy). These differences were particularly acute for deceptive items that lead to the generation of plausible but incorrect responses. In addition, older adults were less effective at monitoring the correctness of responses, as indicated by their lower gamma correlations and greater calibration error.

The current study used the same approach as Kelley and Sahakyan (2003) in order to examine the role of executive processes in memory accuracy. In particular, participants were tested in the memory monitoring task employed by Kelley and Sahakyan and were also administered a battery of neuropsychological tests designed to measure executive functioning (cf., Pansky, Koriat, Goldsmith, & Pearlman-Avnion, 2002). As noted, executive functions may be important in retrieving accurate information and may also be crucial in monitoring the accuracy of candidate responses (Shimamura, 2000a, 2000b). The paradigm to be used in the current study is well suited for examining these issues as one can separately measure the quantity of correct items available at forced report and the accuracy of responses at free report, which will be a function of memory monitoring. One can then determine whether executive functions are predictive of accuracy in free recall. Much of the prior work that has examined individual differences in memory accuracy has assumed that a direct relationship exists between individual differences in executive function and memory accuracy (e.g., Butler et al., in press; McDaniel et al., 1999). The current study makes the novel contribution of examining the role of memory monitoring and indirect effects of executive function in the attainment of accuracy. Specifically, we assessed whether a relationship exists between executive functions and memory accuracy in free report that is mediated by memory monitoring and/or by memory quantity in forced report, independent of a direct relationship between executive functions and memory accuracy.

Two other types of individual difference measures were taken. First, measures of speed of processing were collected (Salthouse, 1996). Salthouse has suggested that age-related differences in cognition are largely the result of a more basic deficit in processing efficacy. Thus, analyses in the current study examined speed of processing in addition to examining executive function correlates of memory accuracy and memory monitoring. Second, working memory was measured using the operation span (OSPAN) task (Turner & Engle, 1989). OSPAN was originally included as an additional measure of executive functioning. Kane and Engle (2002) have suggested that working memory span scores are indicative of a general ability to control attention and maintain information in the face of interference, and therefore tap into a basic component of executive function. However, to preview, results of a factor analysis showed that OSPAN did not load on a factor comprised of several other executive functioning measures. Thus, OSPAN was included as a separate measure of working memory span to determine its contribution to memory accuracy, separate from the contribution of other components of executive function.

Method

Participants

Participants were 50 undergraduates at Florida State University who received course credit for their participation, and 50 older adults aged 65 and over who were recruited from lists of university alumni and through notices published in local newspapers (characteristics of the sample are presented in Table 1). Older adults were paid $10 an hour for their participation. All participants reported good or excellent health and normal or corrected vision. In addition, participants were screened for dementia using the Mini-Mental Status Exam (MMSE; Folstein, Folstein, & McHugh, 1975). All participants achieved scores of 27 or above (i.e., normal cognitive function) and were included in the current study.

Materials

The materials used in the current experiment consisted of 75 word pairs, one third of which were related filler items (e.g., “morning-evening”) and two thirds of which were unrelated word pairs. Half of the unrelated word pairs (*deceptive items*) had potentially interfering competitors that were not only semantically related to the cue word but shared the same first two letters and last letter as the target word (e.g., the “nurse-dollar” pair had the interfering competitor, “doctor,” and the test cue “nurse-do____ r”). The other half of the word pairs consisted of unrelated *control items* that have no such interfering competitors (e.g., “clock-dollar,” with the test cue “clock-do____ r).

Table 1
Means (and *SD*) of demographic characteristics of the participants

Variable	Younger adults	Older adults
N	50	50
Age	19.64 (1.19)	71.84 (5.40)
% Females	56	60
Years of education	14.21 (1.20)	16.74 (2.01)
MMSE[a]	29.48 (.86)	28.72 (1.07)
NAART-35[b]	13.96 (5.16)	22.90 (7.09)

[a] Mini-Mental State Examination (MMSE) score represent the number of points earned out of a possible total of 30.

[b] North American Adult Reading Test-35 (Uttl, 2002) scores represent the total number of correctly pronounced irregular words out of a total of 35 words.

All study items were taken from Kato (1985) and Kelley and Sahakyan (2003). The study list consisted of 60 pairs of words, distributed equally across control items, unrelated deceptive items, and related fillers. As well, a study buffer consisting of one related pair (“month-year”) and one unrelated pair (“turkey-opera”) were included in the study list to serve as practice items at test. The test list consisted of all 60 test items and five new items of each type (i.e., control, unrelated deceptive, and related filler items), for a total of 75 items. Test items were counterbalanced for old-new status and for presentation as control or deceptive items. A 3-item practice test list was also included to familiarize the participant with the test procedure. The practice test list consisted of two pairs of items presented as the study buffer and an additional related pair (“bride-gr__ m”).

Measures

Participants were administered several measures designed to assess executive function, working memory, speed of processing, and vocabulary. These measures are as follows.

Speed of processing measures

The Digit-Symbol Substitution Task (DSST) and the Number Comparison task (Salthouse & Babcock, 1991) were used as measures of speed of processing, along with the Trail Making Test (TMT; Reitan, 1958). The TMT was originally administered as a test of executive function, but loaded on the speed measure rather than the executive function measure in the factor analysis to be reported. For the Digit-Symbol substitution task, the number of correct symbols written from a code table in 90s was recorded as the dependent measure. For the Number Comparison task, the number of correct identifications of pairs of numbers as the same or different (minus the number of incorrect responses) given in two, 90s trials, was recorded as the dependent measure. For the TMT, the time to connect a series of encircled numbers in ascending order (Part A), the time to connect a series of alternating numbers and letters in ascending order (Part B), and the difference in time to complete Parts A and B were recorded as dependent measures.

Executive function measures

The tests of executive function were chosen because they are common measures of executive function and/or because they have been proven to be sensitive to frontal lobe lesions. The computerized version of the WCST (Heaton, Chelune, Talley, Kay, & Curtiss, 1993) was administered, and the two most commonly used measures, number of categories achieved and number of perseverative errors committed, were recorded as dependent measures. The Controlled Oral Word Association test (COWA; or FAS-Test; Benton & Hamsher, 1976) was also administered with the total number of unique words produced across three, 90s trials for the letters *F*, *A*, and *S* recorded as the dependent measure.

Working memory span measure

Working memory capacity was measured using the OSPAN task developed by Turner and Engle (1989). The participant's score was recorded as the total number of words recalled from those trials in which all words were recalled in the correct order.

Procedure

After providing informed consent participants filled out a demographic survey. They were then administered the MMSE. Following the MMSE, participants completed Parts A and B of the Trail Making Test and a vocabulary measure (see Table 1). Next, participants were tested in the memory monitoring task that is the primary focus of the current study. All study and test items were presented via computer using Micro Experimental Laboratory software (Schneider, 1990). Study pairs were presented at an 8 s rate and participants were instructed to remember the pairs for a later test. Words pairs were presented in the center of a computer screen in a large font.

Following a 5-min filler task, participants moved on to the test phase. The cued recall test took place in three stages that occurred consecutively for each test item before the next test item was presented. In the first stage, participants were instructed to use the cue (e.g., nurse do ____ r) to either retrieve an item that was studied or, failing that, to generate an item that fit the cue. This procedure applied to both old and new cues. Next, once the participant either recalled or guessed a target item they were asked to assess the likelihood that the target item was studied using a 0–100% confidence scale. They were informed that a confidence rating of 100% would correspond to a 100% likelihood that the target was the correct answer whereas a confidence rating of 0% would correspond to a 0% likelihood that the target was the correct answer. Participants were encouraged to use the full range of the confidence scale. Following their confidence judgment, participants completed the free report stage, with the test cue remaining on the screen. At free report, participants were instructed to either report whether the item was studied, say "new" if the test cue was new, or "pass" if they were unsure whether the item was old or new. Prior to the test, participants were informed that it would contain some new cues and that they should avoid guessing during the free report phase. Participants were also given instructions to be accurate. This was accomplished by telling participants that they would earn 10 points for each correct answer and would be penalized 10 points for each incorrect answer. They were also instructed that passing would result in no award or penalty. Each stage of the memory monitoring test was self-paced.

Upon completion of the memory monitoring test, participants were given a rest break of 10–20 min. Following the break, participants were administered, in order, the WCST, FAS-Test, DSST, Number Comparison task, and the OSPAN task. Upon completion of these tasks, participants were debriefed and thanked for their participation. The experiment took approximately $1\frac{1}{2}$–3 h to complete.

Results

Analyses for the current study can be broadly categorized into analyses of performance on the memory monitoring task (i.e., memory performance and monitoring performance), performance on the various tests administered, and path models of accuracy. Each of these components of the data analysis will be considered in turn. In the interest of brevity, analyses of memory performance and memory monitoring have been condensed (more detailed analyses are available upon request). Additionally, performance on new items was excluded from all analyses reported. Unless otherwise noted the alpha level for all statistical tests was set to $p < .05$.

Table 2
Means (and *SD*) of quantity and accuracy scores for the free and forced report condition by age and item type

Age group	Item type	Report option		
		Forced	Free	
		Quantity and accuracy	Quantity	Accuracy
Younger	Control	.60 (.18)	.50 (.20)	.81 (.16)
	Deceptive	.42 (.23)	.42 (.23)	.59 (.27)
Older	Control	.53 (.19)	.45 (.22)	.77 (.23)
	Deceptive	.31 (.19)	.29 (.18)	.44 (.26)

Memory performance

As seen in Table 2, young adults generated a greater quantity of correct responses in forced report than did older adults, and also attained a higher level of memory accuracy when they exercised the option of free report. This was confirmed in a 2 (Age: young, old) × 2 (Item Type: control, deceptive) × 2 (Report Option: forced, free) mixed-factor analysis of variance (ANOVA) on memory accuracy scores, with a main effect of age present, $F(1,98) = 6.82$, $MSe = .90$, $d = .53$. Accuracy was higher for all participants at free report ($M = .68$) than at forced report ($M = .47$), $F(1,98) = 259.61$, $MSe = 3.51$, $d = 3.26$, and was also much higher for control items ($M = .68$) than for deceptive items ($M = .44$), $F(1,98) = 168.90$, $MSe = 5.60$, $d = 2.63$. A triple interaction of age, item type, and report option was also present, $F(1,98) = 4.96$, $MSe = .04$, $d = .45$. This reflects the fact that young adults made largely equivalent gains in memory accuracy for control and deceptive items from forced to free report. In contrast, whereas older adults made gains in accuracy for control items that were somewhat larger than those made by young adults, older adults made a less substantial gain in memory accuracy for deceptive items.

There was some trade-off in quantity and accuracy as participants moved from forced to free report (see Table 2). A mixed model ANOVA on the quantity of correct answers revealed that quantity was diminished from forced ($M = .47$) to free report ($M = .43$), $F(1,98) = 67.03$, $MSe = .15$, $d = 1.65$. This finding is qualified by a significant item type × report option interaction, $F(1,98) = 35.73$, $MSe = .06$, $d = 1.21$. Specifically, participants demonstrated greater losses in quantity from forced to free report for control items with no losses evident for deceptive items. Younger adults reported a greater quantity of correct answers compared to older adults, [$F(1,98) = 7.79$, $MSe = 1.03$, $d = .56$], with a trend also evident for older adults to exhibit slightly greater losses in quantity than young adults from forced to free report. However, the interaction of age and report option was only marginally significant, $F(1,98) = 3.45$, $MSe = .01$, $p = .07$, $d = .38$. Thus, overall, participants demonstrated considerable gains in accuracy from forced to free report, at the expense of a moderate loss in quantity correct. Gains in accuracy from forced to free report may reflect the efficacy of monitoring processes, which will be examined next.

Table 3
Memory monitoring data by age and item type

Measure	Item type	Younger adults	Older adults
Calibration error	Control	.18 (.13)	.18 (.14)
	Deceptive	.33 (.21)	.38 (.21)
γ-Correct[a]	Control	.82 (.31)	.78 (.30)
	Deceptive	.76 (.32)	.66 (.53)
ANDI[b]	Control	.56 (.33)	.50 (.34)
	Deceptive	.38 (.35)	.33 (.35)
γ-Response[c]	Control	.92 (.29)	.94 (.10)
	Deceptive	.83 (.30)	.89 (.32)
P_{RC}[d]	Control	.56 (.29)	.63 (.31)
	Deceptive	.61 (.26)	.57 (.31)

[a] Monitoring resolution.
[b] Adjusted Normalized Discrimination Index.
[c] Relationship between confidence and decision to respond.
[d] Response criterion.

Memory monitoring

Participants provided an immediate confidence rating for each item given during forced report.[1] Confidence ratings were then grouped into 12 levels (e.g., 0, .01–.10, .11–.20, .21–.30,…,.91–.99, 1.0). Using these confidence ratings, memory monitoring effectiveness was quantified in terms of calibration and monitoring resolution (see Table 3). *Calibration* refers to the absolute correspondence between assessed levels of confidence and the actual probability that a participant's response is correct (e.g., if perfectly calibrated, items given an assessed confidence of 50% would be correct 50% of the time, and so forth). In contrast, *monitoring resolution* provides a measure of relative correspondence by quan-

---

[1] For 1% of the items, participants changed their response between forced and free report. These items were excluded from the monitoring data reported.

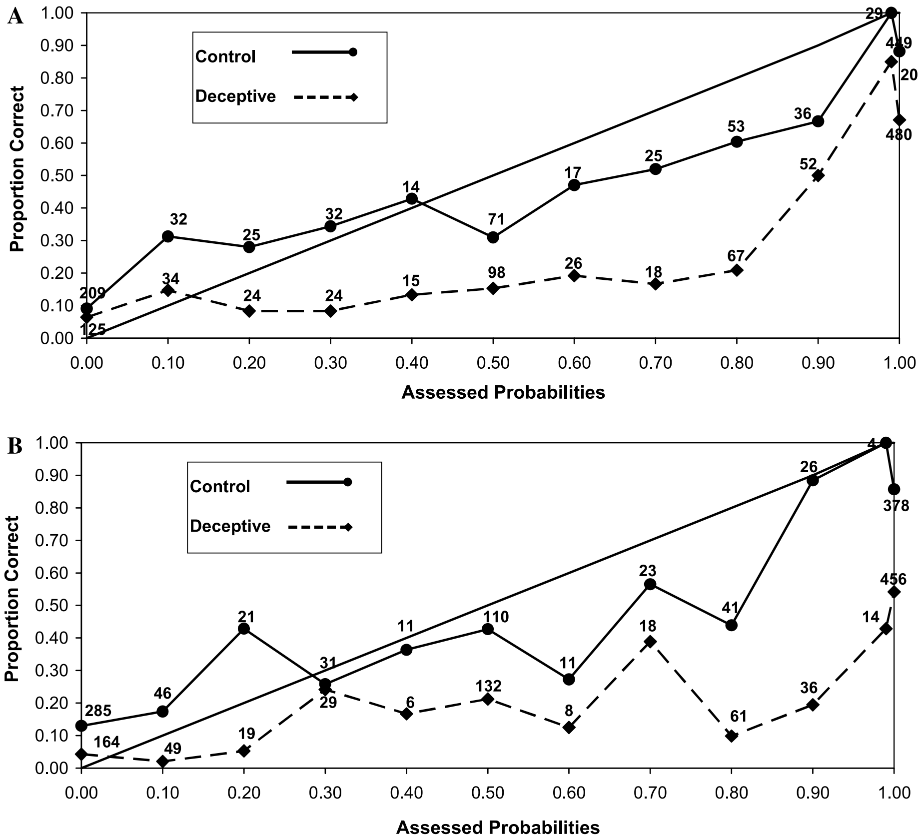

Fig. 1. Calibration curves for deceptive and control items for younger adults (A) and older adults (B). The diagonal line represents perfect calibration while the numbers at each data point correspond to the frequency of responses for each particular level of confidence.

tifying the degree to which assessed confidence distinguishes between correct and incorrect responses.

Fig. 1 displays calibration curves for old (panel A) and young (panel B) participants for control and deceptive items. Both young and older adults were well calibrated for control items. However, calibration on deceptive items was significantly worse for both age groups. For example, young adults reported an average assessed confidence on deceptive items of .70 while their actual proportion correct was .42. Older adults likewise demonstrated considerable overconfidence on deceptive items, with an average assessed confidence of .66 whereas their actual proportion correct was .31. Calibration error scores (the weighted mean of the absolute difference between the assessed probability and the actual proportion correct for each category) were calculated for all participants. A 2 (Age: young, old) × 2 (Item Type: control, deceptive) mixed-factor ANOVA on calibration error scores confirmed that calibration error was greater for deceptive items ($M = .36$) than for control items ($M = .18$), $F(1,98) = 64.07$, $MSe = 1.48$, $d = 1.62$. However, neither the main effect of age, $F < 1$, $d = .20$, nor the interaction of age and item type was significant, $F(1,98) = 1.45$, $MSe = .03$, $p = .23$, $d = .23$.[2] Further inspection of these data revealed that differences in calibration between control and deceptive items did not result from differences in polarization (i.e., instances of using extreme confidence categories of 0 and 100%). Specifically, young and old adults used the categories of 0 and 100% for control items, 66 and 67% of the time, respectively. The pattern is similar for

---

[2] The relationship between confidence and accuracy can also be examined using a global measure of the overall difference between assessed confidence and proportion correct for each item type. Older adults exhibited more overconfidence for deceptive items ($M = .34$) than for control items ($M = .02$). Young adults exhibited the same pattern for deceptive ($M = .28$) and control ($M = .04$) items. Analyses confirmed that participants exhibited greater overconfidence on deceptive items than control items [$F(1,98) = 200.02$, $MSe = 4.00$, $d = 2.86$], but age differences were not apparent, $F < 1$, $d = .15$. However, an age × item type interaction was present [$F(1,98) = 4.43$, $MSe = .09$, $d = .43$], as older adults were more overconfident on deceptive items than younger adults but demonstrated equal overconfidence on control items.

deceptive items, as participants showed polarized responding 62 and 63% of the time, respectively.

Monitoring resolution was assessed using γ correlations (Nelson, 1984) between confidence and memory accuracy at forced report.[3] A mixed-model ANOVA revealed that resolution was marginally better for control items ($M = .81$) than for deceptive items ($M = .71$), $F(1,96) = 3.61$, $MSe = .47$, $p = .06$, $d = .39$. γ correlations did not differ between young and old adults, $F(1,96) = 1.07$, $MSe = .17$, $d = .21$, nor did age interact with item type, $F(1,96) < 1$, $d = .14$. Thus, while monitoring resolution was poorest for deceptive items, age differences were not apparent.

The Adjusted Normalized Discrimination Index (ANDI) was also calculated as a second measure of monitoring resolution (Yaniv, Yates, & Smith, 1991) and ANDI scores were analyzed in a mixed-model ANOVA. Results showed that participants had significantly higher ANDI scores for control items ($M = .53$) than for deceptive items ($M = .36$), $F(1,98) = 16.92$, $MSe = 1.53$, $d = .83$. ANDI scores did not differ on the basis of age [$F(1,98) = 1.20$, $MSe = .19$, $p = .25$, $d = .22$], nor did age interact with item type, $F < 1$, $d = .01$. It must be noted that both these data and data for γ correlations differ from that reported by Kelley and Sahakyan (2003). Specifically, while younger adults' performance was identical between the two studies, older adults exhibited substantially poorer monitoring performance for deceptive items in Kelley and Sahakyan's study. The reason for this discrepancy is unclear but, on a speculative level, may reflect differences between the groups of older adults tested.

Control processes

Control processes influence accuracy by determining whether a response is volunteered or withheld (see Table 3). Overall, there was a strong relationship between the decision to volunteer a response and confidence ratings, particularly for control items. A 2 (Item Type: control, deceptive) × 2 (Age: young, old) mixed-factor ANOVA confirmed that γ correlations between confidence and the decision to respond did not differ by age $F(1,90) = 2.71$, $MSe = .21$, $p = .10$, $d = .35$, or item type, $F(1,90) = 2.54$, $MSe = .21$, $p = .11$, $d = .32$.[4] The interaction of item type and age was not significant, $F < 1$, $d = .07$.

Response criterion estimates (P_{RC}) were also estimated for participants as that confidence value which maximized the number of hits and correct rejections (see Koriat & Goldsmith, 1996). These data were examined in a mixed-model ANOVA. Results showed that response criterion did not differ on the basis of age, $F < 1$, $d = .07$, or item type, $F < 1$, $d = .03$. However, an age × item type interaction was present, $F(1, 97) = 4.86$, $MSe = .19$, $d = .45$. Specifically, older adults exhibited a slightly more stringent criterion for volunteering responses for control items in comparison to deceptive items. Young adults, in contrast, demonstrated the opposite pattern, as they maintained a stricter criterion for deceptive items than for control items.

Discussion

Overall, participants made substantial gains in accuracy from forced to free report and showed striking differences in performance for control versus deceptive items. In particular, participants reported a lower quantity of correct responses on deceptive items in comparison to control items and, as well, did not achieve comparable levels of memory accuracy at free report. Older adults reported a lower quantity of correct responses at forced report and achieved lower levels of memory accuracy at free report, relative to young adults. A lower quantity of correct responses available at forced report does not necessarily lead to diminished accuracy. That is, monitoring processes may operate to achieve accuracy, such that participants only volunteer those responses that are correct. However, participants were persistently overconfident for deceptive items and demonstrated significantly poorer monitoring resolution on such items, though monitoring did not differ on the basis of age.

Psychometric tests: Performance, correlations, and factor loadings

Performance on each of the executive, speed, and memory tests administered is summarized in Table 4. Young adults performed better on each of the tasks administered than older adults with the exception of the verbal fluency task (FAS). This is not completely unexpected, as age differences are not always apparent on the FAS (for a review, see Spreen & Strauss, 1998). Correlations between performance on the various tests administered, accuracy, and monitoring measures are presented in Table 5.

Scores on the Trails A–B difference measure, OSPAN, and each of the other executive and speed measures were subjected to principle components factor analysis. Because the oblique solution resulted in only

---

[3] Occasionally, γ correlations could not be calculated, as in the case when responses were either completely correct or completely incorrect. This occurred for deceptive items for one young participant and one older participant and these data were excluded from the present analysis.

[4] In some cases, γ correlations could not be calculated, as in instances when participants either did not withhold or volunteer any responses. This occurred in seven cases for older participants for deceptive items and once for older participants for control items. In addition, this occurred once for young participants for deceptive items.

Table 4
Means (and *SD*) of scores from executive functioning and speed measures for older and younger adults

Task	Younger adults	Older adults	*F*(1,98)	*d*[a]
Trails making test				
Version A time (s)	22.74 (9.57)	36.28 (11.53)	40.84**	1.29
Version B time (s)	53.50 (25.82)	79.41 (27.36)	23.71**	.98
Difference	30.77 (25.77)	43.13 (20.97)	6.92*	.53
Wisconsin Card Sorting Test				
Categories achieved	5.98 (.14)	4.36 (2.01)	32.39**	1.15
Perseverative errors	7.34 (3.65)	17.18 (12.59)	28.18**	1.07
Verbal fluency[b]	42.74 (11.34)	44.67 (10.93)	.19	.09
Operation span[c]	16.60 (7.23)	12.52 (6.57)	8.72**	.60
Digit symbol substitution[d]	69.52 (9.15)	49.14 (11.74)	93.71**	1.96
Number comparison[e]	54.22 (11.31)	41.08 (11.10)	34.39**	1.18

[a] Effect size.

[b] Scores represent the total number of unique words produced in 1 min for each of the three letters tested (*F*, *A*, and *S*).

[c] Scores represent the total number of words produced in their entirety and in the correct sequence out of a total of 42 words presented.

[d] Scores represent the number of correct responses made in 90 s.

[e] Scores represent the total number of correct responses made in two sessions of 90 s.

* $p < .05$.

** $p < .01$.

a moderate correlation ($r = .28$) the orthogonal solution was accepted. Two factors with eigenvalues exceeding 1.0 were obtained (see Table 6 for a summary). The two speed tests and the difference in Trails A and Trails B performance loaded on the first factor, accounting for 32% of the total variance. The number of categories achieved on the WCST, the number of perseverative errors committed on the WCST, and the number of unique words produced on the FAS-test loaded on the second factor, accounting for an additional 26% of the total variance. The first factor was labeled the *speed composite*. While this label is certainly applicable in the context of the two speed measures it is not conventionally applied to the Trail Making Test. However, the Trail Making Test is a timed task, in common with the speed measures, and optimal performance on the task is predicated on completing it as fast as possible for each portion. This, in conjunction with its strong loading on factor 1, provides sufficient justification to include the difference in times on Parts A and B of the Trail Making Test in the speed composite. The second factor, composed of scores from the WCST and the FAS test, was labeled the *executive composite*. Performance on the OSPAN measure did not load with either factor, and its influence on memory accuracy will be reported separately. Factor scores from both composite measures were used as predictors of individual differences.

A composite measure was also derived for memory monitoring by subjecting ANDI scores and calibration error data for control and deceptive items to principle components factor analysis. Data from gamma correlations were not included as they were somewhat skewed and often subject to ceiling effects such that, on average, participants attained the maximum score (1.0) approximately 35% of the time. In addition, for reasons of exposition that were important for path analyses to be reported, calibration error scores were converted to "calibration goodness" scores by subtraction of calibration error for each participant from 1.

The oblique solution for the factor analysis of monitoring measures resulted in a moderate correlation ($r = .31$) and thus the orthogonal solution was accepted (see Table 7). Two factors with eigenvalues exceeding 1.0 were obtained. All measures of monitoring for deceptive items loaded on the first factor, accounting for 42% of the total variance. Likewise, all measures of monitoring for control items loaded on the second factor, accounting for 41% of the total variance. Therefore, the first factor was labeled the *deceptive monitoring composite* and the second factor the *control monitoring composite*. Weightings based on each factor were used to derive monitoring scores for each type of item.

It must be noted that the monitoring composites derived included measures of both absolute and relative accuracy. Although measures of absolute accuracy, such as calibration error scores, often tap different aspects of monitoring than do measures of relative accuracy (e.g., γ, ANDI), they are highly correlated in this data set. In addition, factor analyses that included only monitoring data (i.e., calibration, ANDI) for either deceptive or control items yielded one factor, suggesting that absolute and relative measures captured similar aspects of performance. Such a finding is reasonable given the paradigm used. For example, the challenge for monitoring deceptive items is that incorrect responses draw high levels of confidence. Hence, deceptive candidate responses

Table 5
Correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Age	1.0																		
2. Trails A–B difference	.28	1.0																	
3. WCST-categories	−.52	−.30	1.0																
4. WCST-persev errors	.47	−.34	−.80	1.0															
5. Verbal fluency (FAS)	(.02)	(−.02)	(.14)	(−.17)	1.0														
6. OSPAN	−.31	(−.18)	.26	−.26	(.17)	1.0													
7. DSST	−.71	−.39	.41	−.41	(.15)	.36	1.0												
8. Number comparison	−.51	−.43	.34	−.35	(.07)	.24	.70	1.0											
9. Forced Quan-Control	(−.20)	(−.12)	(.19)	−.28	.33	.30	(.20)	(.16)	1.0										
10. Forced Quan-Deceptive	−.26	(−.12)	.31	−.24	.29	.22	.26	(.19)	.64	1.0									
11. Free Accuracy-Control	(−.12)	(−.16)	.29	−.30	.29	.32	.15	.06	.69	.48	1.0								
12. Free Accuracy-Deceptive	−.30	(−.13)	.32	−.25	.30	.30	.33	.26	.59	.87	.56	1.0							
13. γ Correct-Control	(−.07)	(−.11)	(.14)	(−.01)	(.02)	(.11)	(.09)	(.07)	(.16)	.21	.41	.26	1.0						
14. γ Correct-Deceptive	(−.12)	(.13)	(.08)	(−.04)	(−.03)	(.16)	(.18)	(.01)	(.14)	.23	(.04)	.34	(.10)	1.0					
15. Calibration Error-Control	(.02)	(−.04)	(−.15)	(.07)	(−.02)	(−.17)	(−.02)	(.06)	−.25	−.22	−.27	−.27	.23	(−.05)	1.0				
16. Calibration Error-Deceptive	(.16)	(.04)	−.26	.22	−.31	−.25	−.26	(−.12)	−.45	−.67	−.40	−.80	(−.17)	−.47	.27	1.0			
17. ANDI-Control	(−.09)	(.01)	(.13)	(−.06)	(.11)	(.19)	(.09)	(.03)	.26	.26	.40	.28	.57	(.01)	−.64	−.24	1.0		
18. ANDI-Deceptive	(−.12)	(.02)	(.12)	(−.07)	(.19)	(.20)	(.07)	(−.10)	.35	.44	.39	.60	(.07)	.36	−.28	−.69	.23	1.0	
19. γ Response-Control	(.10)	(−.01)	(−.01)	(.00)	(.15)	−.21	(−.06)	(.01)	(−.13)	(−.15)	(.09)	(−.09)	.60	(−.08)	(.07)	(.06)	(.16)	(−.10)	1.0
20. γ Response-Deceptive	(.09)	(−.08)	(.11)	(.03)	.22	(.14)	(.07)	(−.05)	(−.01)	(.06)	(.07)	(.14)	(.10)	(.10)	(−.09)	−.28	(.07)	(.16)	(−.03)

Note. With the exception of those in parentheses, all correlations were significant at $p < .05$ or better. WCST, Wisconsin Card Sorting Test; DSST, Digit-Symbol Substitution Task; Quan, Quantity; Decep, Deceptive; ANDI, Adjusted Normalized Discrimination Index.

Table 6
Loading patterns from the orthogonal rotation of the factor analysis of executive scores

Test	Factor 1	Factor 2
Trail Making test: A–B difference	−.70	−.08
Digit symbol substitution	.78	.29
Number comparison	.84	.12
WCST-categories	.38	.74
WCST-perseverative errors	−.39	−.75
Verbal fluency (FAS Test)	−.21	.66
Operation span	.28	.45
Eigenvalue	3.01	1.08
Variance proportion	.32	.26
Correlation		
1. Age	−.61	−.31
2. Factor 1	—	.00

Table 7
Loading patterns from the orthogonal rotation of the factor analysis of monitoring scores

Test	Factor 1	Factor 2
ANDI-Deceptive	.91	.13
Calibration-Deceptive	.91	−.23
ANDI-Control	.11	.90
Calibration-Control	.17	.89
Eigenvalue	2.18	1.16
Variance Proportion	.42	.41
Correlation		
1. Age	−.14	−.03
2. Factor 1	—	.00

will produce high levels of calibration error, and also lower measures of relative accuracy, leading to the strong associations observed.

Path analyses: What accounts for individual differences in memory accuracy at free report?

Path analyses were conducted separately for memory accuracy on deceptive items and memory accuracy on control items in order to determine the contribution of age, executive function, memory, and monitoring to the achievement of accuracy at free report. Each model assumed that age had direct paths to executive function, the quantity of correct items retrieved, and memory accuracy. In addition, each model assumed that executive function, quantity of correct items retrieved, and memory monitoring had direct effects on memory accuracy. However, the models also permitted one to examine whether age and executive function had indirect effects that were mediated by memory retrieval and memory monitoring and whether the quantity of correct items retrieved was mediated by memory monitoring. Preliminary regression analyses indicated that the speed composite measure did not explain any unique variability beyond that explained by age and thus was not included in the path analyses reported.

Fig. 2 depicts the model examined for deceptive items. Overall, the model explained approximately 60% of the variability in memory accuracy at the free report stage for deceptive items (see Table 8 for a summary). Consistent with Kelley and Sahakyan (2003), quantity correct at forced report played a large role in individual

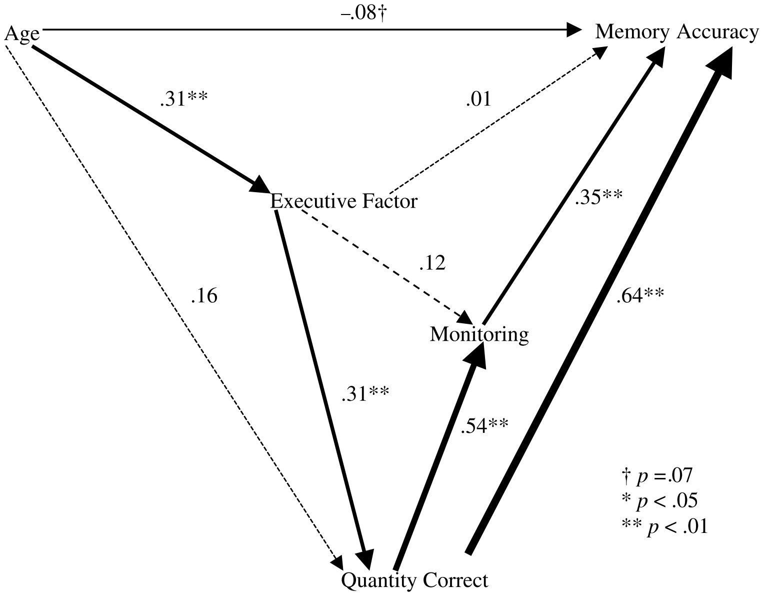

Fig. 2. Path model of memory accuracy at free report for deceptive items. Coefficients for each path represent standardized β weights.

Table 8
Direct and indirect effects for memory accuracy for deceptive items (A) and control items (B)

Predictor	Causal effects		
	Direct	Indirect	Total
(A)			
Age	−.08	−.23	−.31
Executive factor	.01	.30	.31
Quantity correct	.64	.19	.83
Monitoring	.35		.35
(B)			
Age	.07	−.19	.12
Executive factor	.20	.22	.42
Quantity correct	.59	.03	.62
Monitoring	.16		.16

Note. Direct effects reflect the standardized β coefficient for the direct path from a given predictor variable to the dependent variable. Indirect effects were estimated by multiplying those path coefficients that connect a predictor variable to the dependent variable.

differences in free report accuracy for deceptive items. There was a strong direct relationship between the quantity correct on deceptive items at forced report and accuracy at free report, as well as a smaller effect of quantity correct at forced report on accuracy at free report that was mediated by monitoring. Critical to our test of how executive functioning produces individual differences in memory accuracy at free report, we found that the effect of executive functions on free report accuracy was almost entirely mediated by its relationship with quantity correct at forced report. Finally, age had a moderate relationship to accuracy that was largely mediated by its effects on executive processes. This relationship, in turn, had a cascade of effects through executive functions on the quantity of correct items retrieved at forced report and the ability to monitor the correctness of these items.

Memory accuracy at free report for control items was examined in a model identical to that used for accuracy on deceptive items (see Fig. 3). The model explained approximately 31% of the variability in memory accuracy at free report for control items. As in the model for deceptive items, a strong relationship existed between the quantity of correct items retrieved at forced report and ultimate accuracy at free report. However, the indirect effect of quantity correct in forced report that was mediated by memory monitoring was much smaller than was the case for deceptive accuracy. The role of executive functions on free report accuracy was again mediated by its relationship to quantity correct at forced report, but in addition there was a second, direct (albeit weak) path from executive functions to free report accuracy. The relationship between age and memory accuracy in free report was smaller for control items than for deceptive items (as reflected in the overall effect of age in the two models) and that effect was primarily mediated by effects on executive processes. Thus, these data indicate that the executive factor largely affected memory accuracy in free report through indirect effects on quantity correct and not on monitoring. In addition, monitoring was a far weaker predictor of the level of accuracy in free report than was the quantity of correct

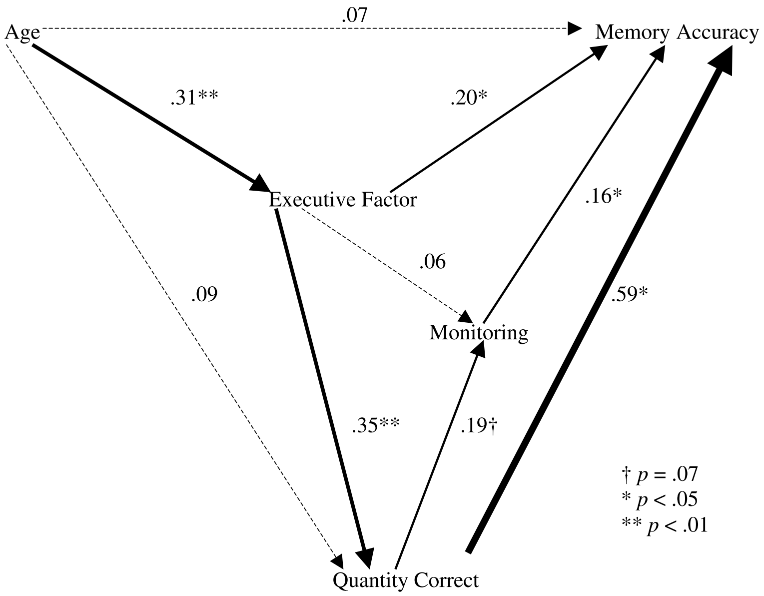

Fig. 3. Path model of memory accuracy at free report for control items. Coefficients for each path represent standardized β weights.

items retrieved in forced report, although monitoring did make a significant, independent contribution to accuracy.

General discussion

We examined the role of age and executive processes in memory accuracy across a sample of young and older adults using an individual differences approach. Age-related differences in memory accuracy were apparent, as well as age-related differences in executive functioning. Most importantly, we found support for two of the major assumptions tested. First, these data indicate that executive functions affect memory accuracy at free report by way of their contribution to the quantity of correct candidate responses available at forced report. Second, memory monitoring partially mediated the effects of quantity correct on accuracy in free report for deceptive items, but not control items. However, we found no support for the assumption that executive functions directly influence the effectiveness of memory monitoring.

The current study was inspired by previous work (Butler et al., in press; McDaniel et al., 1999) highlighting the relationship between executive functioning and age-related changes in memory accuracy. Our work extends their findings to cued recall, and identifies mediators of the executive function-accuracy relationship, namely quantity retrieved in forced report and memory monitoring. However, we also extended their work in that we measured individual differences in executive processing in young adults as well as older adults, and found that the relationship to memory accuracy holds for both groups. For example, younger adults in the top compared to bottom quartile of the young adult distribution of the executive function factor scored 25% points higher on free report accuracy for control items, and 30% points higher on free report accuracy for deceptive items. These data are comparable to the differences for older adults in their top and bottom quartile, and comparable to the differences reported by Butler et al. (in press).

Executive functions and memory performance

The relation between executive functions and the quantity of correct answers retrieved at forced report in the current study fits with data from frontal patient populations suggesting that encoding and retrieval processes that are reliant on executive processes (Moscovitch & Melo, 1997; Norman & Schacter, 1997; Parkin, 1997). The current method of separating performance under forced report from performance under free report permits us to trace the relationship between executive functioning and free report accuracy in our participants to the forced report stage, distinguishing retrieval from monitoring and control processes. How might executive functioning contribute to encoding and retrieval? One possibility is that people who score high on executive function measures either choose better encoding strategies than people lower on executive function measures, or implement strategies more effectively (cf. Bryan, Luszcz, & Pointer, 1999). For example, we recently found that young adults scoring low on measures of executive function (as measured by OSPAN) were more likely to encode a list of words for a test of free recall using rote repetition. In contrast, those with high OSPAN scores used deeper and more relational encoding strategies (Cokely, Kelley, Gilchrist, & Myerson, 2004). Retrieval processes such as elaborating on cues to generate candidate responses (Burgess & Shallice, 1996; see also Jacoby, Shimizu, Velanova, & Rhodes, in press) may also be a more prevalent or effective strategy for people scoring high on measures of executive function.

The specific measure of executive function used in the current study was based on a single factor composed of scores on the WCST and FAS test. However, this may reflect only one aspect of executive function. For example, Miyake et al.'s (2000) analysis of subcomponents of executive functioning argued for three separate executive functions: mental set shifting, information updating and monitoring, and inhibition of prepotent responses (see also Fisk & Sharp, 2004). They reported that WCST performance was most strongly related to mental set shifting. Troyer, Moscovitch, and Winocur (1997) have suggested that a similar function may underlie performance on the verbal fluency test (i.e., the FAS test). Specifically, Troyer et al. proposed that performance on the FAS test requires people to shift to new phonemic cues once they exhaust a cluster such as "things that begin with 'ar'."

The notion that executive function is not a unitary construct is supported by the finding that OSPAN did not load on a composite derived from performance on the other executive measures. In Miyake et al.'s (2000) analysis, OSPAN was most strongly related to updating. Given that OSPAN may reflect a unique subcomponent of executive function, we examined it as a predictor of memory quantity in forced report and memory accuracy in free report in regression analyses. Age was entered first, followed by scores on the OSPAN task, which was followed by the executive component (see Table 9). OSPAN scores explained significant, unique variability in performance on forced report quantity and free report accuracy, but primarily for control items rather than deceptive items. In addition, even after controlling for OSPAN scores, the executive composite measure accounted for a significant amount of variability in performance. Future research relating executive functions to the

Table 9
Proportion of variability accounted for by the executive composite and regression coefficients after controlling for age and OSPAN score for quantity and accuracy measures

Measure	Age alone		After OSPAN score		After executive composite	
	R^2	β	Change in R^2	β	Change in R^2	β
Forced quantity						
Control	.04†	−.20	.07*	.27	.06**	.29
Deceptive	.07**	−.26	.02	.15	.07**	.30
Free accuracy						
Control	.01	−.12	.09**	.31	.10**	.36
Deceptive	.09**	−.30	.05*	.22	.06*	.27

Note. All regression estimates reported are standardized.
* $p < .05$.
** $p < .01$.
† $p = .05$.

attainment of memory accuracy is needed to specify how component processes such as those specified by Miyake et al. contribute to memory encoding, retrieval, and monitoring.

The role of monitoring

Results from the current study showed that memory monitoring processes partially mediated the relationship between quantity retrieved in forced report and accuracy in free report, although those effects were only evident for deceptive items. This likely occurred because monitoring of deceptive items was a far more demanding task than that required for control items (cf. Mather, Johnson, & De Leonardis, 1999), and far more critical in the attainment of accuracy. For example, when accurate retrieval fails for deceptive items at forced report, participants may nonetheless fluently generate the incorrect response, and use the fluent "retrieval" of the deceptive response as a primary basis for mistakenly inferring that the candidate response is a memory. Any gains in accuracy at free report for deceptive items will thus depend on discerning those candidate responses that are correct versus those that are plausible, but incorrect responses (e.g., "doctor" in the case of the cue, "nurse-do____ r"). In contrast, control items (e.g., "clock-do____ r") were far less likely to elicit plausible alternatives and thus easier to monitor.

The role of monitoring may, in fact, be underestimated in this paradigm if participants engage in some monitoring prior to choosing a response at the forced report stage (Koriat & Goldsmith, 1996). Using protocol analysis in this paradigm with younger adults, we have found that participants occasionally engage in monitoring during the forced response stage, as they generate the deceptive (wrong) answer, reject it, and continue attempting to retrieve the correct answer. If there are individual differences in monitoring during forced report, that could produce an overestimation of the role of retrieval in determining memory accuracy in free report and an underestimation of the role of monitoring. More fine-grained methods are needed to capture the possible interplay of retrieval and monitoring processes (cf., Burgess & Shallice, 1996).

Such issues may also be relevant for other measures of memory. For example, although the current study has focused primarily on the role of the frontal lobe in memory vis a vis executive measures, processes localized in the medial temporal region undoubtedly contribute to memory accuracy. Medial temporal lesions have long been associated with amnesia (e.g., Scoville & Milner, 1957) and scores on measures of medial–temporal function are predictive of recollection (Davidson & Glisky, 2002). However, many tests of medial temporal function, such as the California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 1987) and Logical Memory I. of the Wechsler Memory Scale-Revised (Wechsler, 1987), are free report tests. Given this, performance includes monitoring and control processes as well as strict measures of what has been encoded and can be retrieved. In Koriat and Goldsmith's (1996) framework this is akin to measuring accuracy at free report. Therefore, scores derived from medial temporal measures will necessarily reflect the efficacy of monitoring processes, contaminating them as a "pure" measure of memory. The current study did not include a medial temporal battery of tests but instead used memory quantity on items during forced report. In this manner, the contribution of the amount of memory available could be distinguished from the contribution of memory monitoring.

Summary and conclusion

In conclusion, results from the current study indicate that measures of executive functioning are predictive of memory accuracy. While previous studies have examined this relationship based on a free report accuracy measure (e.g., Butler et al., in press; McDaniel et al., 1999) the current study permitted a separate assessment of executive function effects on what is retrieved versus the monitoring of what is retrieved. Results showed that executive processes had their effect on memory accuracy primarily through the quantity of items retrieved at forced report. In turn, quantity correct was a strong predictor of accuracy but was partially mediated by its effect on memory monitoring. Thus, both the accuracy of retrieved responses and memory monitoring must be taken into consideration when individual differences in memory accuracy are examined.

References

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology A, 49*, 5–28.

Bartlett, J. C., Halpern, A. R., & Dowling, W. J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Memory & Cognition, 23*, 531–546.

Benton, K. L., & Hamsher, K. (1976). *Multilingual aphasia examination manual*. Iowa City: University of Iowa.

Bryan, J., Luszcz, M. A., & Pointer, S. (1999). Executive function and processing resources as predictors of adult age differences in the implementation of encoding strategies. *Aging, Neuropsychology, and Cognition, 6*, 273–287.

Burgess, P. W., & Shallice, T. (1996). Confabulation and the control of recollection. *Memory, 4*, 359–411.

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Roediger, H. L., III, Price, & A. L. (in press). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review*.

Chalfonte, B., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24*, 403–416.

Cokely, E. T., Kelley, C. M., Gilchrist, A. L., & Myerson, C. (2004). Nov. Individual differences in working memory and interference from part-set cues. Poster presented at the 45th Annual Meeting of the Psychonomic Society. Minneapolis, MN.

Crawford, J. R., Bryan, J., Luszcz, M. A., Obonsawin, M. C., & Stewart, L. (2000). The executive decline hypothesis of cognitive aging: Do executive deficits qualify as differential deficits and do they mediate age-related memory decline?. *Aging, Neuropsychology, and Cognition, 7*, 9–31.

Davidson, P. S. R., & Glisky, E. (2002). Neuropsychological correlates of recollection and familiarity in normal aging. *Cognitive, Affective, and Behavioral Neuroscience, 2*, 174–186.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California verbal learning test*. New York: Psychological Corporation.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition, 9*, 288–307.

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology, 7*, 874–890.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). 'Mini-Mental State': A practical method for grading the cognitive states of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9*, 229–235.

Glisky, E. L., Rubin, S. R., & Davidson, P. S. R. (2001). Source memory in older adults: An encoding or retrieval problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1131–1146.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Odessa, FL: Psychological Assessment Resources.

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127*, 251–268.

Houx, P. J., Jolles, J., & Vreeling, F. W. (1993). Stroop interference: Aging effects assessed with the Stroop Color-Word Test. *Experimental Aging Research, 19*, 209–224.

Howard, D. V. (1980). Category norms: A comparison of the Battig and Montague (1960) with the response of adults between the ages of 20 and 80. *Journal of Gerontology, 35*, 225–231.

Jacoby, L. L. (1999a). Deceiving the elderly: Effects of accessibility bias in cued-recall performance. *Cognitive Neuropsychology, 16*, 417–436.

Jacoby, L. L. (1999b). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 3–22.

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (in press). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language*.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology, 17*, 3–11.

Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12*, 352–361.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 637–671.

Kato, T. (1985). Semantic-memory sources of episodic retrieval failure. *Memory & Cognition, 13*, 442–452.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48*, 704–721.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123*, 297–316.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–537.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experiment Psychology: Human Learning and Memory, 4*, 19–31.

Mather, M., Johnson, M. K., & De Leonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology, 16*, 437–458.

McDaniel, M. A., Glisky, E. L., Rubin, S. R., Guynn, M. J., & Routhieaux, B. C. (1999). Prospective memory: A neuropsychological study. *Neuropsychology, 13*, 103–110.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Moscovitch, M., & Melo, B. (1997). Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia, 35*, 1017–1034.

Moscovitch, M., & Winocur, G. (1995). Frontal lobes, memory, and aging. *Annals of the New York Academy of Sciences, 769*, 119–150.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–113.

Norman, K. A., & Schacter, D. L. (1996). Implicit memory, explicit memory, and false recollection: A cognitive neuroscience perspective. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 229–257). Mahwah, NJ: Erlbaum.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.

Owens, J., Bower, G. H., & Black, J. B. (1979). The "soap opera" effect in story recall. *Memory & Cognition, 7*, 185–191.

Pansky, A., Koriat, A., Goldsmith, M., & Pearlman-Avnion, S. (2002). Memory accuracy and distortion in old age: Cognitive, metacognitive, and neurocognitive determinants. Poster presented at the 30th Anniversary Conference of the National Institute for Psychobiology in Israel. Jerusalem, Israel.

Parkin, A. J. (1997). The neuropsychology of false memory. *Learning and Individual Differences, 9*, 341–357.

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7*, 290–298.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. Salthouse (Eds.), *The handbook of aging and cognition*. Mahwah, NJ: Erlbaum.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indication of organic brain damage. *Perceptual and Motor Skills, 8*, 271–276.

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin Card Sorting Test: A meta-analytic review. *Psychology and Aging, 19*, 482–494.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language, 35*, 76–100.

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

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403–428.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763–776.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*, 11–21.

Schneider, W. (1990). *MEL user's guide: Computer techniques for real time experimentation*. Pittsburgh: Psychology Software Tools.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114*, 727–741.

Shimamura, A. P. (1996). The role of the prefrontal cortex in controlling and monitoring memory processes. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 259–274). Mahwah, NJ: Erlbaum.

Shimamura, A. P. (2000a). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*, 207–218.

Shimamura, A. P. (2000b). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition, 9*, 313–323.

Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology, 14*, 299–309.

Souchay, C., Isingrini, M., Clarys, D., Taconnat, L., & Eustache, F. (2004). Executive functioning and judgment-of-learning versus feeling-of-knowing in older adults. *Experimental Aging Research, 30*, 47–62.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*, 138–146.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language, 28*, 127–154.

Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology, 24*, 1123–1137.

Vilkki, J., Servo, A., & Surma-aho, O. (1998). Word list learning and prediction of recall after frontal lobe lesions. *Neuropsychology, 12*, 268–277.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., & de Menezes Santos, M. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119–125.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised manual*. New York: Psychological Coroporation.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120*, 272–292.

Whelihan, W. M., & Lesher, E. L. (1985). Neuropsychological changes in frontal function with aging. *Developmental Neuropsychology, 1*, 371–380.

Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin, 110*, 611–617.