



ELSEVIER

Contents lists available at SciVerse ScienceDirect

## Journal of Memory and Language

journal homepage: [www.elsevier.com/locate/jml](http://www.elsevier.com/locate/jml)

## Retrieval effort improves memory and metamemory in the face of misinformation <sup>☆</sup>

John B. Bulevich <sup>\*</sup>, Ayanna K. Thomas

*Department of Psychology, The Richard Stockton College of New Jersey, Galloway, NJ 08205, United States*

*Department of Psychology, Tufts University, Medford, MA, United States*

### ARTICLE INFO

#### Article history:

Received 3 September 2011

revision received 18 December 2011

Available online 21 January 2012

#### Keywords:

Misinformation

Metamemory

Confidence

Aging

Eyewitness memory

Retrieval effort

### ABSTRACT

Retrieval demand, as implemented through test format and retrieval instructions, was varied across two misinformation experiments. Our goal was to examine whether increasing retrieval demand would improve the relationship between confidence and memory performance, and thereby reduce misinformation susceptibility. We hypothesized that improving the relationship between confidence and memory performance would improve controlled processes at retrieval. That is, when confidence and memory performance were well calibrated, participants would be able to withhold incorrect responses if given the opportunity. To examine the relationship between memory retention, confidence, and controlled withholding, we compared older and younger adults' performance on a forced memory test, where participants could not withhold responses, and on a free test, where participants were encouraged to withhold responses. Confidence judgments were collected after forced responding. Retrieval demand was manipulated indirectly through type of test (cued recall vs. recognition) and directly through retrieval instructions. The results demonstrated that increasing retrieval demands improved memory retention, metamemorial monitoring and effective withholding. This was particularly pronounced when participants received misleading information. Finally, older adults required explicit direction to effectively monitor memory and institute successful controlled withholding.

© 2012 Elsevier Inc. All rights reserved.

### Introduction

Consider the possibility of witnessing a crime. You are most likely motivated to report what you have witnessed correctly. However, when interviewed by detectives, you find that in their attempts to ascertain a complete account of the event, they encourage you to guess and persuade you to answer questions even when you are unsure. While questioning, these detectives may also unwittingly introduce information that may bias or distort your recollection. Even the most earnest witness will find it difficult to

achieve accuracy when faced with pressures to provide details, when encouraged to guess, when presented with post-event information, and when discouraged to say "I don't know". The present study examined changes in younger and older adult eyewitness memory accuracy within the misinformation paradigm under conditions where people were forced to respond to every question and under conditions where people were encouraged to exercise control and withhold answers for which they may not be certain. We examined performance under forced-reporting instructions to ascertain memory retention of a witnessed event after the presentation of misleading post-event information. We examined performance under free-reporting instructions to determine whether the effect of misleading post-event information could be counteracted, thereby improving memory accuracy. Finally, we manipulated retrieval demands in order to improve memory retention and accuracy.

<sup>☆</sup> Portions of this research represents a dissertation for a Doctor of Philosophy degree at Washington University in St. Louis for the first author.

<sup>\*</sup> Corresponding author. Address: The Richard Stockton College of New Jersey, Department of Psychology, 101 Vera King Farris Drive, Galloway, NJ 08205, United States.

*E-mail address:* [john.bulevich@stockton.edu](mailto:john.bulevich@stockton.edu) (J.B. Bulevich).

### *Retention, monitoring, and control*

In their classic study, Loftus, Miller, and Burns (1978) demonstrated that exposure to misleading information after witnessing an event reduced accuracy on a later memory test. Variants of this general finding have since been demonstrated in dozens of papers. Further, older adults have been shown to be more susceptible to misinformation than younger adults (Cohen & Faulkner, 1989; Coxon & Valentine, 1997; Holliday et al., 2011; Karpel, Hoyer, & Togli, 2001; Mitchell, Johnson, & Mather, 2003). A primary goal of the present research is to examine processes that underlie age-related increases in misinformation susceptibility, in order to develop techniques to improve overall memory performance. The Quantity-Accuracy Profile (QAP) methodology, developed by Koriat and Goldsmith (1996), allows us to examine the contributions of memory retention disruption, monitoring and/or control deficits, to misinformation susceptibility. Specifically, this model incorporates metacognitive processes into the assessment of memory performance. Retention is measured by the forced-reporting memory test and is the proportion of forced report answers that are correct. Metacognitive monitoring is also measured during forced-reporting through the collection of confidence judgments for each answer. Memory monitoring effectiveness can be captured by monitoring resolution, which allows for an evaluation of whether participants can distinguish correct from incorrect responses. Control is measured during the free-reporting test, which follows forced-report. During this testing phase participants are re-presented with responses given during the forced report phase. Participants are encouraged to exercise control and withhold answers that may be associated with low confidence. By comparing performance during the forced and free reporting stages we can assess both control sensitivity and response criterion. That is, we can determine whether participants volunteered answers that accompany high confidence and determine the confidence point at which answers were volunteered. The former allows for the examination of the effect of monitoring on control. Koriat and Goldsmith's model makes the assumption that monitoring will affect control (see also, Nelson & Narens, 1990) and that individuals will withhold answers associated with low confidence (Goldsmith & Koriat, 2008).

Research suggests that misinformation susceptibility may, in part, be the result of damage or occlusion to the memorial information. For example, the memory impairment hypothesis (Loftus, 1979) states that the memory trace for the original information is permanently altered or destroyed by the misleading post event information. Similarly, the blocking hypothesis (Eakin, Schreiber, & Sergent-Marshall, 2003), suggests that the original information is "blocked" from access by the misleading information. Measures of retention would directly examine these memory-based explanations for misinformation susceptibility, as broadly, they suggest that the original information is not accessible at the time of test.

Alternative to retention explanations, misinformation susceptibility may be a result of erroneous subjective experience that resulted in faulty decisions at retrieval. As one example, in order for participants to effectively withhold

answers, they must be able to successfully inspect and monitor both the quality of evidence for a candidate response and the bases for responding when recollection fails. Regarding the first point, when specific details are recollected, individuals often will determine that the episode was actually experienced. However, when conscious recollection fails, individuals may use alternatives to recollection such as plausibility, familiarity, or accessibility (Jacoby & Hollingshead, 1990; Reder, Wible, & Martin, 1986; Thomas, Bulevich, & Chan, 2010). These alternatives, recruited under forced-responding test conditions, may be less valid than detailed recollection and may lead to memory inaccuracies. In the case of the misinformation effect, participants may respond with the misleading post-event information because that information is easily accessed (i.e., Thomas et al., 2010). When given the opportunity to withhold during free-responding, those answers may be omitted. That is, when given the opportunity to use monitoring and control processes, people may assess the effectiveness of retrieval strategies, and if allowed, may withhold items that were accessed with retrieval strategies judged to be less effective or reliable. By using the QAP methodology, and comparing older to younger adults, we examined the specific cognitive mechanisms that underlie increased age-related susceptibility to misinformation (i.e., Mitchell, Johnson, & Mather, 2003). Further, we developed age-appropriate techniques to counteract the negative influence of misleading post-event information.

### *Age-related monitoring deficits*

Across the varied metacognitive monitoring tasks, older adults have sometimes shown age-related impairments and, in other instances, age-equivalent performance. For example, older adults have been shown to be less accurate than younger adults when making episodic FOK predictions (Perrotin, Isingrini, Souchay, Clarys, & Tacconat, 2006; Souchay, Isingrini, & Espagnet, 2000; Thomas, Bulevich, & Dubois, 2011); however age equivalence has been demonstrated on semantic FOK tasks (Allen-Burge & Storandt, 2000; Butterfield, Nelson, & Peck, 1988; Bäckman & Karlsson, 1985; Lachman, Lachman, & Thronesbery, 1979; Marquie & Huet, 2000). Research suggests that the inconsistency of age-related impairment in monitoring tasks might be due to the specific requirements of the task, with episodic FOK judgment accuracy being more dependent on access to contextual episodic information (Souchay, Moulin, Clarys, Tacconat, & Isingrini, 2007; Thomas et al., 2011).

In the present study, we were interested in the monitoring of the validity of candidate memory responses as expressed by the relationship between confidence judgments and correctness (Lovell & Marsh, 1985; Perfect & Stollery, 1993), within the misinformation paradigm. Previous research has demonstrated that older adults show a pattern of overconfidence when exposed to misleading post-event information (Cohen & Faulkner, 1989; Dodson, Bawa, & Krueger, 2007; Dodson & Krueger, 2006; Jacoby, Bishara, Hessels, & Toth, 2005; Karpel et al., 2001; Mitchell et al., 2003). This pattern of overconfidence may be a result of an accessibility bias. That is, information that is easily accessed, or retrieved more fluently, may result in higher

confidence ratings than information that is less accessible. In situations where misleading information is more accessible, confidence in that retrieved information may be inappropriately high. Further, Kelley and Sahakyan (2003) found that the monitoring deficit demonstrated by older adults resulted in ineffective withholding. In this task, older and younger participants were presented with words pairs that were either weakly related (*dog-rock*) or weakly related and deceptive (*nurse-dollar*). These pairs were described as deceptive because the authors tested participants' memories with the cue *nurse-d\_\_\_\_\_r*. In this situation, the pre-potent response to the cue is *doctor* rather than *dollar*. The pattern of age-related changes in both monitoring and control were intensified for deceptive word pairs (see Rhodes & Kelley, 2005 for similar results). Finally, Pansky, Goldsmith, Koriat, and Pearlman-Avni (2009) also demonstrated age-related monitoring and control deficits in older adults after the presentation of a narrative slide show. In the present study, we attempt to synthesize these different components by examining whether misleading post-event information negatively impacted these age effects in monitoring and control and whether the age deficit could be reduced.

Younger adults have also demonstrated overconfidence in the misinformation paradigm (Weingardt, Leonesio, & Loftus, 1994). Similarly, when young adult participants were asked to make remember/know (R/K) judgments on the criterial test after exposure to misleading post-event information, participants gave "remember" responses to items incorrectly retrieved from the narrative and attributed them to the original event (Roediger, Jacoby, & McDermott, 1996). While there is an ongoing debate as to how to interpret remember/know judgments, there is substantial support for the view that these judgments are at least partially supported by confidence, (Wixted & Stretch, 2004; Yonelinas, 2001; but see Karayianni & Gardiner, 2003 for an alternate viewpoint). Taken together, these results suggest that misleading post-event information negatively impacts the relationship between confidence and memory performance, and over-confidence is greater in older adults as compared to younger adults.

#### *Improving monitoring and control after misinformation*

A second goal of the present research was to examine techniques that could be used to improve metacognitive processes of monitoring and control. Using the QAP methodology, we directly assessed whether techniques designed to enhance performance affected retention, monitoring, and/or control. We tested whether increasing retrieval demands would improve the relationship between memory performance and confidence, thereby improving accuracy on the final free-report test. The influence of retrieval demands on these processes was examined in two ways: (1) comparing cued recall to recognition; (2) providing instructions to participants designed to facilitate the retrieval of additional contextual information before answering questions on the forced-reporting test.

Researchers generally agree that retrieval demands differ between recognition and recall tests of memory, with recall placing larger retrieval demands on the partic-

ipant (Roediger & Guynn, 1996). Whether recall tests are conceptualized as requiring more retrieval effort or increased controlled processing, it is clear that participants engage in different and, most likely, more taxing processes than those engaged when recognition tests are given. In support of our retrieval demand hypothesis, Robinson, Johnson, and Robertson (2000) found that the relationship between confidence and memory performance was better when participants were given a cued recall as compared to a recognition test (see also Robinson & Johnson, 1996, for a similar result). To extend this finding, we examined whether retrieval demand would affect this relationship after exposure to misleading post-event information. Research has also demonstrated that retrieval of contextual information prior to making metacognitive judgments led participants to be more accurate in their predictions (Thomas, Bulevich, & Dubois, 2011). Further, when older and younger participants were instructed to evaluate contextual information before making source monitoring decisions, they improved the accuracy of those decisions (Thomas & Bulevich, 2006). The present experiments examined the role retrieval demands play in reducing general susceptibility to and age-related differences in the misinformation effect.

#### *The present experiments*

The present study examined the contributions of retention, monitoring, and control to misinformation susceptibility in younger and older adults. Both experiments followed the same general methodology. First, participants viewed a 40 min video. Following the video, they read a narrative that included both correct and incorrect information. After a brief delay, participants engaged in their first memory test. They were told to base their answers on what occurred in the video. In the first test, participants were forced to produce or select an answer for each question whether or not they felt that they actually knew the answer. They also assigned each answer a confidence judgment. In the second phase of testing, participants received a copy of the test they completed with the confidence ratings removed. They were told, in this phase, to select only the answers that they actually believed to be correct (i.e., actually occurred in the video).

Retention was measured during the forced reporting test phase. We predicted that memory accuracy would be lower on misleading trials as compared to control trials in both older and younger adults; however, we expected a greater misinformation effect in older than younger adults. Monitoring was also measured during the forced reporting phase, by comparing confidence judgments with memory performance. We predicted that older adults would demonstrate impaired monitoring when compared to younger adults on misleading trials; however, the age differences would be mediated by increasing retrieval demands through type of test and type of instructions. This prediction follows from findings where older adults' feeling of knowing accuracy improved after requiring the retrieval of contextual information related to the target (Thomas, Bulevich, & Dubois, 2011) as well as results that showed that diagnostic instructions improved older adults'

memory performance in a source monitoring task (Thomas & Bulevich, 2006). The age discrepancy in monitoring would result in control deficits, with older adults being less likely to withhold incorrect answers as compared to younger adults.

## Experiment 1

### Method

#### Participants

The participants were 48 younger and 48 older adults. The participants were tested in groups of either older or younger participants and received either course credit or financial compensation (\$10 per hour). The younger adults were Washington University undergraduates and the older adults were recruited from the Washington University and Colby College Older Adult participant pools. The demographic information for the participants for both experiments is displayed in Table 1. Older adults were prescreened, and only those who presented as cognitively healthy, were not on medications known to affect cognitive functioning, and who had normal or corrected-to-normal hearing and vision were tested. The older adults did not differ on level of education or vocabulary score (Salthouse, 1993) from the younger adults.

#### Design

The experimental design was a  $2 \times 2 \times 2$  mixed design. The variables were Age (Older or Younger adults), Type of Test (Cued Recall or Recognition) and Trial Type (Control or Misleading). Age and Type of Test were between subjects variables.

#### Materials

The materials consisted of a 42 min video clip. The video was the first episode of the first season of 24 (20th Century Fox Television, 2001). The video details the events of a government agent attempting to stop a political assassination. Twenty-four critical items were selected from the video as items for the memory test.

To select the 24 critical items, the materials were pilot tested. Fifteen younger adults and fifteen older adults were shown the video clip. They were then given a cued recall test with 52 items. These items were selected to correspond to pieces of information in the video that were clearly visible or audible as well as easily verifiable. The

tests were scored and items were eliminated if over 90% of the respondents answered correctly or less than 10% answered correctly from either group. After these eliminations, 24 items remained which served as the critical items.

For the misinformation segment of the experiment, a written narrative was generated from the video. Twenty-four critical items were divided into two groups; 12 items served as *control* items and were not included in the narrative at all and 12 items provided *misleading* information based on what occurred in the video. The misleading information in the narrative contradicted details from the video (i.e., person's car described as a different color). Item type was counterbalanced across participants.

#### Procedure

Participants were tested in groups of four to ten. In the first phase of the experiment, participants were instructed to watch the video carefully and to pay attention to details because their memories were going to be tested later. Participants watched the video projected on a Sony LCD projector from an approximate distance of twelve to eighteen feet. The image measured 112 inches on the diagonal. None of the participants had previously seen the video.<sup>1</sup> After the video concluded, participants were given a copy of the narrative, which was approximately 1100 words. They were told that it was a written narrative of the video they had watched and were given 10 min to read it. Participants were not given any additional information regarding the author or accuracy of the narrative. After 10 min, the narratives were collected and participants completed a short demographics form, a synonym vocabulary test (Salthouse, 1993), and a payment voucher, if necessary. These forms took approximately 15 min to complete.

Afterwards, participants began the first phase of testing. Thirty younger and 30 older adults received a four alternative recognition test, and 18 younger and 18 older adults received a cued recall test. A set of participants had to be excluded from analysis for the cued recall test group due to experimenter error. Participants were told to base all answers on what they remembered from the video and to disregard the narrative. Testing was divided into two phases, a forced-responding phase and a free responding phase. During the forced responding phase, participants were required to answer **every** question. Specifically, for the recognition test, participants were instructed that they must *select* an answer for every question and not to leave any blank even if they must guess. For the cued recall test, participants were instructed to produce an answer for every question. Participants were also instructed to provide a confidence rating for every answer on a scale of 25% to 100% with 25% representing chance performance for the recognition test. They were instructed to provide a confidence rating for every answer for the cued recall test on a scale of 0–100%. The questions were presented in the

**Table 1**  
Demographic information for both experiments (means and standard deviation).

	Age	Years of education	Vocabulary score
<i>Experiment 1</i>			
Younger	20.1 (1.5)	14.1 (0.9)	14.2 (1.8)
Older	74.2 (4.1)	15.2 (3.1)	14.8 (4.2)
<i>Experiment 2</i>			
Younger	19.5 (0.7)	13.7 (1.4)	14.2 (2.2)
Older	73.2 (3.6)	14.6 (5.6)	14.5 (2.7)

<sup>1</sup> None of the participants had seen the particular episode of the program used in the experiment. However, based on pilot testing, previous viewing of the episode did not significantly affect participants' responses. In addition, participants from both age groups reported being engaged during the presentation of the episode based on a post-experimental questionnaire.

order that they occurred in the video and the participants were told to take as long as needed for the test. The mean time to complete the test was 17 min.

Once completed, the tests were collected and the experimenter removed the confidence ratings from the test. The free-responding testing phase then began. Specifically, participants were presented with the answers they provided during forced-responding and were instructed to circle only the items that they actually believed to be correct. Confidence ratings were removed so that participants would not arbitrarily pick a number (i.e., 75%) and select all answers given that rating or higher. After they completed this stage of testing, the participants were debriefed and dismissed.

## Results

The primary goal of this experiment was to examine memory retention, monitoring of retention, and the relationship between monitoring and control within the misinformation paradigm. To measure retention, we examined performance associated with forced-responding. Monitoring of retention was measured by examining resolution through Somers' D correlations.<sup>2</sup> The Somers' D correlation (Pannu & Kaszniak, 2005) is a measure of association between two variables that ranges from  $-1$  to  $+1$ . Values close to an absolute value of 1 indicate a strong relationship between the two variables and values close to 0 indicate little or no relationship between the variables. Somers' D is an asymmetric extension of gamma that differs only in the inclusion of number of pairs not tied on the independent variable. To examine the relationship between monitoring and control we focused on gains in accuracy between forced and free responding as well as control sensitivity. To preview our findings, older adults demonstrated a greater misinformation effect, as measured by the difference between performance on control and misleading trials on the forced-responding test, than younger adults. The misinformation effect was smaller on the cued recall test as compared to the recognition test. Monitoring of retention, as measured by Somers' D correlations, was better on the cued recall test as compared to recognition. Control was impaired on misleading trials as compared to control trials; however it was better when participants were given a cued recall test as compared to a recognition test. Finally, participants were able to use monitoring and control processes to improve their accuracy between forced and free responding.

### Retention and monitoring of retention

By examining the proportion of forced-responding answers that were correct, we were able to assess overall

<sup>2</sup> We chose to include resolution (the item by item relationship between confidence and memory performance) as our measure of monitoring effectiveness. Calibration (the aggregate relationship between confidence and memory performance) was a less interesting and informative measure of monitoring effectiveness in the context of our experiments. When calibration served as the dependent measure, the effects that we found (overconfidence for misleading items and increased overconfidence for older adults) replicated previous research (Dodson & Krueger, 2006; Weingardt, Toland, & Loftus, 1994) and only served to make the exposition of the results more difficult.

retention. As displayed in Table 2, we found evidence for the misinformation effect in both older and younger adults. Further, older adults were more susceptible to misleading post-event information than younger adults. To confirm this pattern, we performed a 2 (Trial Type: Control, Misleading)  $\times$  2 (Group: Younger, Older)  $\times$  2 (Test: Cued Recall, Recognition) mixed ANOVA and found main effects for Trial Type,  $F(1,92) = 63.96$ ,  $p < .01$ , Group,  $F(1,92) = 72.53$ ,  $p < .01$ , and Test,  $F(1,92) = 27.49$ ,  $p < .01$ . Performance was worse on misleading trials ( $M = .44$ ) as compared to control trials ( $M = .61$ ), demonstrating the misinformation effect. Younger adults ( $M = .64$ ) demonstrated greater overall retention than older adults ( $M = .42$ ). In addition, retention was greater when participants were given a recognition test ( $M = .60$ ) as compared to a cued recall test ( $M = .46$ ). Finally, the interaction between Test and Trial Type was significant,  $F(1,92) = 10.77$ ,  $p < .01$ . The discrepancy in retention between control trials and misleading trials was greater when participants

**Table 2**  
Measures of memory and monitoring broken down by age and type of test for Experiment 1.

	Control		Misleading	
	M	SD	M	SD
<i>Younger</i>				
Cued recall				
Forced report				
Quantity	.63	(.12)	.55	(.21)
Accuracy	.63	(.12)	.55	(.21)
Monitoring	.65	(.37)	.64	(.35)
Free report				
Quantity	.48	(.20)	.49	(.21)
Accuracy	.81	(.22)	.86	(.17)
Prc	.86	(.15)	.81	(.18)
Recognition				
Forced report				
Quantity	.79	(.12)	.58	(.15)
Accuracy	.79	(.12)	.58	(.15)
Monitoring	.54	(.35)	.25	(.31)
Free report				
Quantity	.58	(.17)	.50	(.19)
Accuracy	.96	(.07)	.80	(.24)
Prc	.88	(.14)	.90	(.10)
<i>Older</i>				
Cued recall				
Forced report				
Quantity	.39	(.17)	.26	(.14)
Accuracy	.39	(.17)	.26	(.14)
Monitoring	.64	(.29)	.46	(.39)
Free Report				
Quantity	.28	(.13)	.21	(.13)
Accuracy	.63	(.27)	.51	(.14)
Prc	.84	(.23)	.76	(.25)
Recognition				
Forced report				
Quantity	.63	(.16)	.39	(.20)
Accuracy	.63	(.16)	.39	(.20)
Monitoring	.56	(.31)	.09	(.44)
Free report				
Quantity	.46	(.17)	.31	(.18)
Accuracy	.79	(.18)	.45	(.27)
Prc	.84	(.18)	.79	(.18)

were given a recognition test as compared to a cued recall test. That is, the misinformation effect was greater when participants took a recognition test as compared to a cued recall test. No other effects were significant,  $F_s < 1$ .

Resolution, using the Somers' D correlations, was computed to examine monitoring accuracy. A 2 (Trial Type: Control, Misleading)  $\times$  2 (Group: Younger, Older)  $\times$  2 (Test: Cued Recall, Recognition) ANOVA on Somers' D correlations found a main effect for Group,  $F(1,89) = 19.03$ ,  $p < .001$ , and a main effect of Trial Type,  $F(1,89) = 19.41$ ,  $p < .001$ . Younger adults demonstrated better resolution ( $M = .59$ ) than older adults ( $M = .36$ ). Resolution was also better on control trials ( $M = .59$ ) as compared to misleading trials ( $M = .36$ ). In addition, the interaction between Group and Trial Type was significant,  $F(1,89) = 7.73$ ,  $p < .05$ . The resolution difference between older and younger adults was greater on misleading trials as compared to control trials, with older adults demonstrating extremely poor resolution on misleading trials. Finally, the interaction between Trial Type and Test was marginally significant,  $F(1,89) = 2.70$ ,  $p = .09$ . The resolution difference between control and misleading trials was greater when participants were given a recognition test as compared to a cued recall test, with recognition resulting in extremely poor resolution on misleading trials ( $M = .17$ ). No other effects were significant,  $F_s < 1$ .

#### *How monitoring affects control: gains and losses*

We examined whether older and younger adults could use metacognitive processes to improve their accuracy between forced and free responding. Accuracy on the forced responding test was identical to our retention measure, and was computed as the proportion of correct answers out of the total number of questions. Accuracy on the free responding test was the proportion of correct answers out of the total number of answers provided (Kelley & Sahakyan, 2003; Koriat & Goldsmith, 1996). To begin with we conducted a 2 (Responding: Forced, Free)  $\times$  2 (Test: Cued Recall, Recognition)  $\times$  2 (Trial Type: Control, Misleading)  $\times$  2 (Group: Older, Younger) ANOVA to examine gains in accuracy between forced and free responding. Gains were demonstrated through a main effect found for Responding,  $F(1,92) = 259.07$ ,  $p < .001$ . That is, performance was improved when participants were given the option to withhold responses ( $M = .72$ ) as compared to when forced to respond ( $M = .53$ ). Responding significantly interacted with Test,  $F(1,92) = 16.00$ ,  $p < .001$ . The gains in performance were greater when participants were given a cued recall test as compared to a recognition test. Further, as with the measure for retention, we found main effects for Group,  $F(1,92) = 80.54$ ,  $p < .001$ , and Test,  $F(1,92) = 11.82$ ,  $p < .001$ . Performance scores were higher for younger adults as compared to older adults. In addition, performance scores were higher when participants were given a recognition test as compared to a cued recall test. With regard to Trial Type, consistent with the retention measure, performance scores were significantly higher for control trials as compared to misleading trials,  $F(1,92) = 45.09$ ,  $p < .001$ . Further, as with our measure of retention, Trial Type interacted with Group,  $F(1,92) = 6.11$ ,  $p < .05$  and Test,  $F(1,92) = 14.26$ ,  $p < .001$ .

More important to the present set of analyses, we found that the interaction among Responding, Group, and Test, was marginally significant,  $F(1,92) = 2.84$ ,  $p < .1$ . Both younger and older adults made greater gains in accuracy when given a cued recall test as compared to a recognition test; however the difference between forced report scoring and free report scoring was greater for older adults than younger adults. That is, older adults benefitted more from the cued recall test than younger adults. Further, the interaction among Responding, Trial Type, and Group,  $F(1,92) = 7.16$ ,  $p < .01$  was significant. As can be seen in Table 2, whereas younger adults made gains for both control and misleading trials, older adults made greater gains on control trials as compared to misleading trials. Finally, the interaction among Responding, Trial Type, and Test was marginally significant,  $F(1,92) = 2.79$ ,  $p < .1$ . The greatest gains were made on misleading trials when participants were given a cued recall test. No other effects were significant,  $F_s < 1$ .

If participants cannot perfectly distinguish which of their candidate responses in forced report are correct and which are incorrect, then increases in accuracy in free report may come at the expense of quantity correct. We were primarily interested in how much correct information was lost between forced and free responding, and whether that loss differed as a function of Trial Type. We performed a 2 (Responding: Forced, Free)  $\times$  2 (Test: Cued Recall Recognition)  $\times$  2 (Trial Type: Control, Misleading)  $\times$  2 (Group: Younger, Older) ANOVA on quantity averages under forced and free responding constraints. Quantity was determined by computing the proportion correct out of the total number of questions. Thus even under free responding constraints, where participants were given the option to withhold responses, the total number of questions still determined the average scores.

Under these scoring parameters we did find that participants lost correct information when given the opportunity to withhold. This was demonstrated by the main effect found for Responding,  $F(1,92) = 108.92$ ,  $p < .001$ . Performance scores were higher under forced responding constraints ( $M = .59$ ) as compared to free responding constraints ( $M = .42$ ). We also found a main effect of Trial Type,  $F(1,92) = 33.45$ ,  $p < .01$ . As in previous performance analyses, participants performed better on control trials as compared to misleading trials. Also consistent with previous analyses, we found a main effect of Group,  $F(1,92) = 62.80$ ,  $p < .001$ , and a main effect of Test,  $F(1,92) = 20.14$ ,  $p < .001$ . Younger adults performed better than older adults, and performance was better when participants were given a recognition test as compared to a cued recall test. As it relates to the examination of losses, Responding interacted with Test,  $F(1,92) = 4.34$ ,  $p < .01$ . Losses were greater when participants were given a recognition test as compared to a cued recall test. Responding also interacted with Trial Type,  $F(1,92) = 33.94$ ,  $p < .001$ . Losses were greater on control trials as compared to misleading trials. Finally, the interaction between Trial Type and Test was significant,  $F(1,92) = 7.07$ ,  $p < .005$ . The difference in performance scores between cued recall and recognition was greater on control trials as compared to misleading trials. No other effects were significant,  $F_s < 1$ .

### Control sensitivity and response criteria setting

We compared Somers' D correlations between confidence in each answer and whether or not it was volunteered to assess control sensitivity. Both older ( $M = .83$ ) and younger ( $M = .87$ ) adults demonstrated control sensitivity, indicating that both groups volunteered answers on the free-responding test that were associated with high confidence on the forced-responding test. A 2 (Group: Young, Old)  $\times$  2 (Test: Cued Recall, Recognition)  $\times$  2 (Trial Type: Control, Misleading) ANOVA on these average Somers' D correlations did not find any significant differences or interactions,  $F_s < 1$ .

Response criterion estimates were derived following the procedure developed by Koriat and Goldsmith (1996) for each trial type. They defined response criterion as the value on the confidence scale that determines the volunteering decision rule: Any item receiving a confidence rating equal to or greater than the response criterion is volunteered, while those with confidence ratings below the criterion are withheld. To determine the response criterion for each individual, we treat each probability rating used by a participant as a possible candidate for a response criterion, and then compute the proportion of items conforming to the decision rule. That is, hits are defined as volunteered items that had an assessed probability greater than or equal to the response criterion, and correct rejections are defined as withheld items that had an assessed probability less than the response criterion. The proportion of hits and correct rejections for a possible response criterion is called the fit ratio, and the judgment category with the largest fit ratio is considered to be that participant's response criterion, PRC (see also, Goldsmith & Koriat, 2008; Kelley & Sahakyan, 2003). We conducted a 2 (Trial Type: Control, Misleading)  $\times$  2 (Test: Cued Recall, Recognition)  $\times$  2 (Age: Younger, Older) ANOVA on PRC and found a main effect of Trial Type,  $F(1,92) = 4.54, p < .01$ . Response criterion was set higher for control trials ( $M = .86$ ) as compared to misleading trials ( $M = .82$ ). In addition, we found a main effect of Age,  $F(1,92) = 4.21, p < .05$ . Younger adults ( $M = .86$ ) set a higher response criterion than older adults ( $M = .81$ ).

### Discussion of Experiment 1

Both older and younger adults demonstrated the misinformation effect. That is, retention was poorer on misleading trials as compared to control trials. The misinformation effect was greater in older adults than younger adults. Finally, while retention was greater on the recognition test as compared to the cued recall test, the misinformation effect (the difference in retention between control and misleading trials) was smaller on the cued recall test as compared to the recognition test. Post-event misinformation also negatively affected monitoring information retained in memory. That is, the relationship between monitoring and retention was poorer on misleading trials as compared to control trials. Further, older adults demonstrated a greater monitoring deficit on misleading trials as compared to younger adults. These data are consistent with previous research (Dodson & Krueger, 2006; Robinson et al., 2000), and support the conclusion that older adults'

poorer performance in the misinformation paradigm may be due to inaccurate subjective experiences.

When given the opportunity to withhold responses on the free responding test, both older and younger adults withheld responses that accompanied lower confidence judgments, as demonstrated by the control sensitivity analysis. While gains in accuracy resulted when participants were given the opportunity to withhold responses, these gains were accompanied by a loss of the completeness of the report. Importantly, for both older and younger adults the loss in completeness was greater for control trials as compared to misleading trials. Finally, the impairment in monitoring demonstrated by older adults negatively influenced their ability to improve memory performance when given the opportunity to withhold answers. That is, younger adults demonstrated greater gains in performances when given the opportunity to withhold answers as compared to older adults. This finding may be due in part to the different response criteria set by each age group. That is, older adults did set a lower response criterion when compared to younger adults.

## Experiment 2

We wanted to determine whether retrieval demands, when dissociated from test format, would also lead to improvements in metacognitive accuracy. We hypothesized that when participants were given a cued recall test they would engage in a more detailed and comprehensive search process as compared to when given a recognition test. These search processes should result in both better relative calibration and more effective withholding. In Experiment 1, type of test was confounded with hypothesized search processes. In Experiment 2, we attempted to decouple test format from retrieval demands by using a procedure that would facilitate a more comprehensive search process regardless of type of test. Detailed instructions were given to participants before both the recognition and cued recall tests. These instructions were designed to increase the likelihood that participants would search for relevant perceptual and contextual cues. We predicted that when given these supportive instructions, metacognitive performance would improve even when participants were given a recognition test. That is, the relatively shallow search process engaged with the recognition test would be enhanced with supportive instructions.

### Method

#### Participants

The participants were 60 younger and 53 older adults. The participants were tested in groups of either older or younger participants and received either course credit or financial compensation (\$10 per hour). The younger adults were Tufts University undergraduates and the older adults were recruited from the Tufts University Older Adult participant pool. The demographic information for the participants for both experiments is displayed in Table 1. As in Experiment 1, older adults were prescreened, and only those who presented as cognitively healthy, were not on

medications known to affect cognitive functioning, and who had normal or corrected-to-normal hearing and vision were tested. The older adults did not differ on level of education or vocabulary score (Salthouse, 1993) from the younger adults.

### Design

The experimental design was a  $2 \times 2 \times 2 \times 2$  mixed design. The variables were Group (Older or Younger Adults), Type of Test (Cued Recall or Recognition), Retrieval Support (Supportive Instructions or Standard Instructions), and Trial Type (Control or Misleading). Group, Type of Test, and Retrieval Support were between subjects variables.

### Materials

The materials were identical to those used in the previous experiments.

### Procedure

In the first phase of the experiment, participants were instructed to watch the video carefully and to pay attention to details because their memories were going to be tested later. Participants were seated in front of a computer and watch the video using Windows Media Player. Participants wore headphones. After the video concluded, participants were given a narrative. The narratives used in Experiment 2 were identical to those used in Experiment 1. After 10 min, the narratives were collected and participants completed a short demographics form, a vocabulary test (Salthouse, 1993), and a payment voucher, if necessary.

Afterwards, they began the first phase of testing. The testing phase was similar to the previous experiments, with the exception that Retrieval Support was varied as well. Fifty-three older and younger participants completed the cued recall test. Sixty participants completed the recognition test. Within each test, participants were further divided into two instructions conditions, a “standard instructions” and a “supportive instructions” condition. In all conditions there were 15 younger adults; thus 60 younger adults participated in this experiment. In the “standard instructions” recognition, and the “supportive instructions” recognition conditions, there were 15 older adults respectively. In the “standard instructions” cued recall condition, there were 12 older adults, and in the “supportive instructions” cued recall condition, there were 11 older adults. Three older adults in the standard instructions condition had to be excluded as well as four in the supportive instructions condition due to noncompliance with the instructions. Specifically, they did not provide answers to all items on the forced test.

In the supportive instructions condition, participants were given additional instructions that were designed to increase elaborative search processes on the forced test. Specifically, both older and younger adults were told “when attempting to answer questions we would like you to take your time and carefully evaluate all aspects of the information you retrieve. For example, if you are asked about the color of a person’s shirt, try and determine if you can actually remember the visual image in your head. Or if you are asked about what someone said, can you remember the sound of

their voice when they said it? Remembering contextual information that occurred directly before or after a particular event is also helpful in determining the accuracy of your memory. Memory for these details is a good indicator that your memory is accurate. If you lack this type of information when you try to remember, it may indicate that your memory is not accurate. When answering these questions try to carefully inspect these details.” After these instructions, each question was read to the participants. The experimenter waited for participants to answer and provide a confidence rating before reading the next question. The free report memory test followed in the same manner as described in the previous experiment.

### Results

#### Retention and monitoring of retention

To assess overall retention and the standard misinformation effect we performed a  $2$  (Trial Type: Control, Misleading)  $\times 2$  (Support: Supportive, Standard)  $\times 2$  (Test: Cued Recall, Recognition)  $\times 2$  (Group: Older, Younger) ANOVA. The misinformation effect was confirmed with a significant effect of Trial Type,  $F(1, 105) = 35.51$ ,  $p < .001$ . As can be seen in Table 3, performance was worse on misleading trials ( $M = .49$ ) as compared to control trials ( $M = .62$ ). A main effect of Group was also found,  $F(1, 105) = 104.44$ ,  $p < .001$ . Older adults demonstrated poorer retention as compared to younger adults. Importantly, the interaction between Trial Type and Group was significant,  $F(1, 105) = 3.99$ ,  $p < .05$ . As in Experiment 1, older adults demonstrated a greater misinformation effect as compared to younger adults.

As it relates to retrieval demands, we found significant effects for both Test,  $F(1, 105) = 52.02$ ,  $p < .001$ , and Support,  $F(1, 105) = 12.23$ ,  $p < .001$ . Overall retention was poorer when participants took a cued recall test ( $M = .47$ ) as compared to a recognition test ( $M = .65$ ). Retention was better when participants were given supportive instructions ( $M = .60$ ) as compared to standard instructions ( $M = .51$ ). Finally, the interaction between Trial Type, Group, and Test was significant,  $F(1, 105) = 10.09$ ,  $p < .005$ . Older adults demonstrated a greater misinformation effect when given a recognition test as compared to younger adults; however, the misinformation effect demonstrated under cued recall test constraints did not vary as a function of age.

To examine monitoring effectiveness at retention we performed a  $2$  (Trial Type: Control, Misleading),  $\times 2$  (Support: Supportive Instructions, Standard Instructions)  $\times 2$  (Test: Cued Recall, Recognition)  $\times 2$  (Group: Older, Younger) ANOVA on Somers’ D correlations. To begin with, we found a main effect of Group,  $F(1, 105) = 11.75$ ,  $p < .005$ . Younger adults ( $M = .56$ ) demonstrated better resolution than older adults ( $M = .45$ ). In addition, we found main effects for Support,  $F(1, 105) = 10.98$ ,  $p < .001$ . Resolution was better when participants were given supportive instructions ( $M = .56$ ) as compared to standard instructions ( $M = .45$ ). We also found main effects for Trial Type,  $F(1, 105) = 14.50$ ,  $p < .001$ , and Test,  $F(1, 105) = 8.20$ ,  $p < .005$ . Resolution was better when participants took a cued recall test ( $M = .55$ ) as compared to a recognition test

**Table 3**

Measures of memory and monitoring broken down by age, type of test, and type of instructions for Experiment 2. Standard deviations are in parentheses.

	Standard				Support			
	Control		Misleading		Control		Misleading	
	CR	REC	CR	REC	CR	REC	CR	REC
<i>Younger</i>								
Forced								
Quantity	.61 (.13)	.76 (.18)	.52 (.27)	.68 (.15)	.67 (.11)	.86 (.16)	.51 (.22)	.84 (.16)
Accuracy	.61 (.13)	.76 (.18)	.52 (.27)	.68 (.15)	.67 (.11)	.86 (.16)	.51 (.22)	.84 (.16)
Monitoring	.58 (.26)	.53 (.25)	.61 (.20)	.37 (.27)	.60 (.24)	.60 (.25)	.58 (.24)	.58 (.21)
Free								
Quantity	.52 (.16)	.61 (.14)	.49 (.23)	.48 (.26)	.49 (.20)	.53 (.12)	.46 (.24)	.55 (.26)
Accuracy	.90 (.14)	.91 (.15)	.76 (.24)	.77 (.35)	.85 (.20)	.90 (.12)	.82 (.16)	.93 (.13)
Prc	.82 (.15)	.91 (.13)	.84 (.12)	.91 (.13)	.78 (.16)	.85 (.16)	.83 (.15)	.88 (.12)
<i>Older</i>								
Forced								
Quantity	.33 (.15)	.58 (.17)	.26 (.18)	.35 (.11)	.45 (.15)	.69 (.21)	.38 (.19)	.39 (.14)
Accuracy	.33 (.15)	.58 (.17)	.26 (.18)	.35 (.11)	.45 (.15)	.69 (.21)	.38 (.19)	.39 (.14)
Monitoring	.55 (.23)	.56 (.35)	.37 (.22)	.03 (.21)	.54 (.27)	.61 (.23)	.55 (.19)	.38 (.22)
Free								
Quantity	.25 (.23)	.40 (.14)	.15 (.18)	.26 (.21)	.31 (.17)	.35 (.20)	.36 (.25)	.29 (.18)
Accuracy	.62 (.38)	.86 (.17)	.45 (.29)	.36 (.18)	.70 (.23)	.81 (.28)	.65 (.26)	.53 (.18)
Prc	.71 (.28)	.73 (.25)	.87 (.15)	.84 (.15)	.84 (.14)	.84 (.18)	.83 (.18)	.84 (.15)

( $M = .46$ ). Resolution was better for control trials ( $M = .57$ ) as compared to misleading trials ( $M = .43$ ). As it relates to retrieval demands, the interactions between Test and Support,  $F(1, 105) = 4.02$ ,  $p < .05$ , and Support and Trial Type,  $F(1, 105) = 3.99$ ,  $p < .05$  were both significant. Supportive instructions had a greater impact on the recognition test as compared to the cued recall test. Similarly, supportive instructions had a greater impact on misleading trials as compared to control trials. Interactions between Trial Type and Age,  $F(1, 105) = 7.34$ ,  $p < .05$ , and Trial Type and Test,  $F(1, 105) = 7.45$ ,  $p < .05$ , were also significant. Younger and older adults demonstrated comparable monitoring effectiveness on control trials; however, older adults were significantly impaired on misleading trials as compared to younger adults. Finally, the difference in monitoring effectiveness between control and misleading trials was greater when participants took a recognition test as compared to a cued recall test.

#### How monitoring affects control: gains and losses

Performance under the constraints of forced and free responding was examined to assess gains and losses. Gains were assessed by comparing forced responding performance, or retention, with accuracy when participants were given the opportunity to withhold responses. Importantly, accuracy under free responding constraints was obtained by dividing the total number correct by the total number answered. To assess gains, we performed a 2 (Trial Type: Control, Misleading)  $\times$  2 (Responding: Forced, Free)  $\times$  2 (Group: Older, Younger)  $\times$  2 (Test: Cued Recall, Recognition)  $\times$  2 (Support: Supportive Instructions, Standard Instructions) ANOVA. As expected, we found a main effect of Trial Type,  $F(1, 105) = 46.39$ ,  $p < .001$ . Performance scores were higher for control trials as compared to misleading trials. We also found a main effect of Responding,  $F(1, 105) = 84.55$ ,  $p < .001$ . Replicating Experiment 1, performance scores were higher under free responding scoring

constraints as compared to forced responding scoring constraints. Also replicating Experiment 1, we found main effects for Group,  $F(1, 105) = 129.21$ ,  $p < .001$ , and Test,  $F(1, 105) = 17.89$ ,  $p < .001$ . Younger adults performed better than older adults, and performance was better when participants were given a recognition test as compared to a cued recall test. Finally, we found a main effect of Support,  $F(1, 105) = 8.30$ ,  $p < .001$ . Performance was better when participants were given supportive instructions ( $M = .67$ ) as compared to standard instructions ( $M = .61$ ).

Consistent with Experiment 1, Trial Type interacted with Age,  $F(1, 105) = 8.49$ ,  $p < .005$ . That is, older adults demonstrated a greater discrepancy between control and misleading trials as compared to younger adults. In addition, Trial Type interacted with Support,  $F(1, 105) = 6.34$ ,  $p < .05$ . Supportive instructions resulted in greater improvement on misleading trials as compared to control trials. Finally, the interaction among Trial Type, Test, and Age was significant,  $F(1, 105) = 9.87$ ,  $p < .005$ . When participants were given a cued recall test, both older and younger adults demonstrated comparable differences between control and misleading trials. However, when given a recognition test, the difference between control and misleading trials was significantly smaller for younger adults as compared to older adults.

Consistent with Experiment 1, we found that Responding interacted with Test,  $F(1, 105) = 84.53$ ,  $p < .001$ . When forced to respond, performance scores were greater when participants were given a recognition test as compared to a cued recall test; however, when given the opportunity to withhold, there was no difference in performance scores between cued recall and recognition testing. The interaction between Responding, Trial, and Support was also significant,  $F(1, 105) = 14.53$ ,  $p < .001$ . In the standard instructions group, participants demonstrated gains between forced and free responding on both control and misleading trials; however, the gains were greater on control

trials as compared to misleading trials. On the other hand, in the supportive instructions group, the gains were relatively small on control trials ( $M_{gain} = .02$ ), yet quite large on misleading trials ( $M_{gain} = .20$ ). No other effects were significant,  $F$ 's < 1.

In order to examine losses, as with gains, we compared performance under the constraints of forced and free responding; however, accuracy when participants were given the opportunity to withhold responses was measured by dividing the total number correct by the total number of questions. We conducted a 2 (Trial Type: Control, Misleading)  $\times$  2 (Responding: Forced, Free)  $\times$  2 (Group: Older, Younger)  $\times$  2 (Test: Cued Recall, Recognition)  $\times$  2 (Support: Supportive Instructions, Standard Instructions) ANOVA. To begin with, we found main effects for Trial Type,  $F(1, 105) = 31.05$ ,  $p < .001$ , Responding,  $F(1, 105) = 74.38$ ,  $p < .001$ , Age,  $F(1, 105) = 125.58$ ,  $p < .001$ , and Test,  $F(1, 105) = 27.36$ ,  $p < .001$ . All of these main effects replicate findings from Experiment 1. Specifically, we found that performance scores were higher on control trials as compared to misleading trials, and that younger adults performed better than older adults. In addition, performance scores were higher when participants were given a recognition test as compared to a cued recall test. We also found that performance scores were higher under forced responding scoring criteria ( $M = .56$ ) as compared to free responding loss criteria ( $M = .44$ ). This finding demonstrates that there was a loss of accurate information between forced and free responding.

The interaction among Trial Type, Group, and Test was also significant,  $F(1, 105) = 5.83$ ,  $p < .05$ . When given a cued recall test, the difference between control and misleading trials was similar between younger and older adults; however, older adults demonstrated a much larger difference between control and misleading trials when given a recognition test as compared to younger adults. When losses were directly assessed by examining the variable of Responding, we found that Responding interacted with both Support,  $F(1, 105) = 7.23$ ,  $p < .01$ , and Test,  $F(1, 105) = 14.22$ ,  $p < .001$ . Under the constraints of forced report scoring, performance scores were higher when participants were given supportive instructions as compared to standard instructions. However, type of instructions did not influence responding when performance on the free responding test was examined. That is, when given the opportunity to withhold, participants demonstrated losses in accuracy; however, those losses were not influenced by type of instructions. Regarding the interaction between Responding and Test, while performance was generally better when participants were given a recognition test as compared to a cued recall test, the discrepancy between cued recall and recognition was small under free reporting constraints as compared to force reporting constraints. The interaction between Trial Type and Responding was also significant,  $F(1, 105) = 4.63$ ,  $p < .05$ . The discrepancy between control and misleading trials was greater under forced reporting constraints as compared to free reporting constraints. Finally, the interaction between Trial Type, Responding, and Support was significant,  $F(1, 105) = 4.03$ ,  $p < .05$ . Under free responding constraints, the difference between control and misleading trials was smaller when

participants were given supportive instructions ( $M_{loss} = .03$ ) as compared to standard instructions ( $M_{loss} = .09$ ). However, the difference between control and misleading trials was not affected by instructions under forced reporting constraints.

#### Control sensitivity and response criteria setting

We compared Somers' D correlations between confidence in each answer and whether or not it was volunteered to assess control sensitivity. A 2 (Group: Young, Old)  $\times$  2 (Test: Cued Recall, Recognition)  $\times$  2 (Support: Supportive Instructions, Standard Instructions) ANOVA on average Somers' D correlations associated with misleading trials did not find any significant differences or interactions,  $F$ 's < 1. A 2 (Group: Young, Old)  $\times$  2 (Test: Cued Recall, Recognition)  $\times$  2 (Support: Supportive Instructions, Standard Instructions) ANOVA on PRC, found a main effect of Age,  $F(1, 92) = 3.85$ ,  $p < .05$ . Younger adults ( $M = .85$ ) set a higher response criterion than older adults ( $M = .81$ ). No other effects were significant.

#### Discussion of Experiment 2

When retrieval demands were directly instantiated through supportive instructions, older adults were able to effectively use those instructions on both cued recall and recognition memory tests. That is, they demonstrated improvements in resolution and memory accuracy across both tests of memory. For younger adults, the most significant benefits of supportive instructions were found when they were given a recognition test. In fact, supportive instructions had minimal effect on metamemory monitoring when younger adults were given a cued recall test. These results suggest that younger adults *automatically* engaged a deeper, more effective search of memory when given a test (cued recall) that required those processes for successful metamemorial performance. Older adults needed additional explicit direction in order to improve metamemorial monitoring. When examining the pattern from forced to free reporting, the supportive instructions improved performance for both younger and older adults. In addition, the relative gains associated with the supportive instructions were larger for misleading items compared to control. Older adults also disproportionately benefitted from the support. However, the supportive instructions did not accompany a larger loss of correct information when transitioning from forced to free responding.

#### General discussion

Research has clearly demonstrated that people incorporate suggested, but incorrect, information into their memory for an event (cf. Loftus, Miller, & Burns, 1978; the present experiments). In the present study, we used the misinformation methodology in order to examine the relationship between retention, monitoring, and control for complex episodic events. In addition, the misinformation methodology allowed us to examine whether improvements in memory and metamemory could be made even in cases where participants had to distinguish between

multiple and sometimes conflicting sources of information. To briefly summarize our findings, both groups demonstrated diminished monitoring after misinformation, with older adults performing even worse than younger adults. However, when retrieval demands were increased through test or supportive instructions, monitoring and controlled withholding improved in both age groups. Direct (type of instructions) and indirect (type of test) retrieval demand manipulations had different effects on retention and monitoring. When retrieval demands were increased indirectly through test, we found that in both age groups, monitoring improved but retention decreased. When we increased retrieval demands through instructions, younger adults demonstrated greater improvements on the recognition test as compared to the cued recall test. Supportive instructions positively affected older adults' retention. Supportive instructions also greatly improved older adults' retention, monitoring, and controlled withholding, even for misleading items.

#### *Memory retention is affected by retrieval demands*

We examined memory retention through performance on forced responding tests. The test required that participants respond to every question. This test format strips away significant aspects of the metamemorial processes that are included in many other memory tasks (conveying certainty or choosing not to answer). In the context of both experiments, retrieval demands had clear effects on memory processes. Specifically, the cued recall test reduced retention in both older and younger adults, as compared to the recognition test. However, the pattern was different when retrieval demands were increased via supportive instructions. For younger adults, instructions improved retention on the recognition test. For older adults, supportive instructions improved retention across both tests. We hypothesized that supportive instructions encouraged both younger and older adults to engage in a more effortful search process. The additional aid provided by these instructions may have been unnecessary when younger adults were given a cued recall test. Research suggests that a cued recall test may be more demanding than a recognition test (Roediger & Guynn, 1996); as such, younger adults may have automatically engaged in a more diagnostic search when given a cued recall test.

Our research suggests that older adults may require additional direct support in order to retrieve and use contextual information. This pattern is consistent with those found by Thomas and Bulevich (2006). Specifically, our earlier research demonstrated that when older and younger adults were instructed as to how to effectively differentiate real from imagined memories, they reduced erroneous memories for performed actions in the imagination inflation paradigm. Similarly, Hasher and colleagues (see Iker, Yang, & Hasher, 2008, for an example) demonstrated that, particularly under interference conditions, older adults did not automatically engage in search strategies that would lead to optimal performance when compared to younger adults. The present study suggests that older adults may be able to recruit useful strategies, but require some direction. This is similar to Craik's (1983) assertion

that older adults require environmental support in order to better utilize their available cognitive resources. Environmental support has been utilized as an explanation for improved older adult performance when applied to such diverse manipulations as matching encoding and retrieval contexts (Naveh-Benjamin & Craik, 1995) to requiring source monitoring judgments at retrieval (Hashtroudi, Johnson, & Chrosniak, 1989). The instructions given in the present study were done so at the beginning of the testing session and participants had to flexibly and strategically apply those instructions to each judgment made. After instructions, participants did not receive any additional source monitoring prompts. That is, they were not encouraged on a judgment-by-judgment basis to discriminate between the original event and the narrative. Rather, participants had to remember the initial instructions in order to benefit.

Our supportive instructions are conceptually similar to instructions provided by Lane, Roussel, Starns, Villa, and Alonzo (2008), in which they provided diagnostic information to participants regarding the characteristics of veridical vs. false memories in the Deese–Roediger–McDermott paradigm. Interestingly, these instructions were successful in reducing that type of memory errors in younger but not older adults. Another recent misinformation study (Holliday et al., 2011) demonstrated improvements in older adult performance through a retrieval manipulation. Holliday et al. used the cognitive interview after exposure to misinformation and found that it improved memory performance in both younger and older adults. They argued that their cognitive interview emphasized the retrieval of specific perceptual details of the event rather than the general information from the event. This is similar to our supportive instructions, which emphasize the recollection of contextual information as a way to distinguish between accurate memories and plausible, but suggested ones. Clearly, in certain situations older adults can benefit from instructions; however, the present body of research does not definitively demonstrate the nature of these situations. Across these studies, techniques to support retrieval and materials vary widely. Older adults may not be able to benefit from supportive instructions in situations where to-be-remembered information lacks a rich narrative structure, as is the case in DRM experiments.

#### *Metamemory is affected by retrieval demands*

In these experiments, we examined metamemorial monitoring through resolution. Resolution is a measure of one's ability, on an item-by-item basis, to distinguish between correct and incorrect answers (Nelson & Narens, 1990). If resolution is good, participants will give high confidence ratings to correctly remembered information and low confidence ratings to incorrectly remembered information. Further, our research rests on the assumption that metamemorial monitoring affects metacognitive control (see Nelson & Narens, 1990, but see Koriat, Ma'ayan, & Nussinson, 2006 for a different perspective). If monitoring affects control then when given the opportunity to withhold responses on the free responding tests participants should withhold those that accompanied low confidence.

To begin with, both older and younger adults demonstrated poor resolution on misleading trials as compared to control trials. Resolution was improved in both age groups when participants were encouraged to engage in a more effortful retrieval process in a cued recall test (see Roediger & Gynn, 1996). Further improvement in resolution was garnered with supportive instructions. Younger adults demonstrated dramatic improvements for misleading items when given supportive instructions before taking a recognition test. Older adults demonstrated improvements on both cued recall and recognition tests after receiving supportive instructions. These results suggest that when given a cued recall test, younger adults may automatically search for and effectively use relevant contextual information when making confidence judgments. However, when given a test that required the retrieval of additional contextual information, older adults may still need accompanying support to improve their monitoring. This is consistent with the finding of smaller age related deficits on recognition test compared to recall tests ( Craik & McDowd, 1987) as well as the general assertion that older adults are particularly impaired in self-initiated controlled processing and disproportionately benefit from environmental support (Craik, 1983).

In addition, it is important to note that the supportive instructions did not merely lead to a general criterion shift in terms of volunteering information. Specifically, after the supportive instructions, participants did not simply become more conservative. If that were the case, we would not have observed the differential effects of the supportive instructions on control as compared to misleading trials. Borrowing from assumptions made by the source monitoring framework, we hypothesize that misleading post-event information may lead to a possible conflict at test between two sources of information (Johnson, Hashtroudi, & Lindsay, 1993). A shallow retrieval process may bias participants towards the most easily accessible information (see also Thomas et al., 2010). The misinformation may be more easily accessed because it is temporally closer to test than the original event. A deeper, more effortful retrieval process, as engendered through test format, would require that participants retrieve more contextual information. That contextual information likely aided monitoring as it aided retention. Research from other metacognitive domains supports this conclusion. For example, Thomas, Bulevich, and Dubois (2011) found that feeling-of-knowing (FOK) judgments were more accurate in predicting future recognition when participants retrieved contextual information as compared to when they did not. Further, in order for older adults to demonstrate improvements in FOK accuracy, contextual information had to be retrieved *before* making the FOK judgment. More recent research demonstrates that accuracy of FOK judgments is also influenced by the amount of correct contextual information retrieved (Thomas, Dubois, & Bulevich, in preparation). This is similar to Robinson et al. (2000) where vividness was the largest predictor of metamemorial accuracy. We would expect vividness to be strongly related to the retrieval of contextual information. Thus, the present findings suggest that the relationship between confidence and memory performance is improved when participants are encouraged to

retrieve contextual information. Contextual information is more likely to be retrieved under the constraints of cued recall as compared to recognition due to the inherent demands of the test.

Finally, because older adults and younger adults differed significantly with regarding to retention, in order to isolate metamemorial differences we compared resolution on misleading trials between a subset of older and younger participants whom had been matched on retention. Even on this subset of older and younger adults, we found that younger adults ( $M = .48$ ) demonstrated significantly better resolution as compared to older adults ( $M = .07$ ),  $t(37) = 3.75$ ,  $p < .005$ . A similar pattern was found in Experiment 2, when we compared 24 older adults ( $M = .35$ ) to 24 younger adults ( $M = .58$ ),  $t(46) = 2.65$ ,  $p < .01$ . These results suggest that older adults' deficits in the misinformation paradigm may be driven by monitoring deficits that cannot be explained through impairment in retention alone.

With regarding to withholding, the present results demonstrate that for both older and younger adults, monitoring influenced withholding. As the control sensitivity data indicate, both groups withheld items that accompanied low confidence. Thus, by improving monitoring resolution, we improved the effectiveness of withholding. Specifically, when retrieval demands were increased older and younger participants were able to take better advantage of the opportunity to withhold responses. As with the effect on monitoring effectiveness, younger participants reaped the greatest reward from the supportive instructions on the recognition test. The supportive instructions improved older adults' controlled processing across both types of test. Older adults were able to improve their performance greatly by withholding items when given supportive instructions compared to when they were not (29% vs. 1%). Further, the gains garnered through support did not accompany greater losses. Taken together, this suggests that the age-deficit in episodic memory tasks may at least partially be due to a failure at retrieval to search for contextual information that would inform their memorial decisions.

## Conclusions and final thoughts

Control of memory and the subjective experiences that accompany it remain a topic of debate and inquiry among psychologists. The present experiments examined how retrieval demands influenced confidence and control over memory in the misinformation paradigm. The experiments demonstrated that older adults' are more susceptible to misinformation as compared to younger adults. We also demonstrated that older adults showed greater metamemorial deficits as compared to younger adults. As compared to younger adults, they were more impaired in regard to monitoring effectiveness, and less likely to improve memory accuracy when given the opportunity to withhold responses. Most importantly, our research demonstrated that the cues derived from the additional retrieval effort engendered by test format and instructions were instrumental in improving metamemorial accuracy. The improvement was greater in older adults than younger adults.

The complex relationship found in the present study between monitoring and control suggests that explanations of the misinformation effect should not neglect metamemorial processes. The present data strongly suggests that the disruption of metamemorial monitoring was a large factor in producing memory errors in this paradigm especially for older adults. However, we were able to improve performance in several different ways. Allowing people to withhold answers compared to forcing them to respond, and increasing retrieval demands both reduced errors. Thus allowing people to exercise control and engage more effortful retrieval processes (either through test format or instructions) appears to improve the relationship between confidence and memory performance.

## Acknowledgments

This research was partially supported by a departmental research grant from Washington University, a National Institute of Health grant, and a Tufts University Faculty Research Award. We would like to thank Roddy Roediger and Pascal Boyer who provided helpful comments and feedback from the genesis through the completion of this research. We would also like to thank Joe Nah, Samia Zahran, and Johanna Seiden for help with data collection and coding.

## References

- 20th Century Fox Television (2001). 24 [DVD]. (Available from 20th Century Fox Home Video, P.O. Box 900, Beverly Hills, CA 90213-0900).
- Allen-Burge, R., & Storandt, M. (2000). Age equivalence in feeling-of-knowing experiences. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, 55B(4), 214–233.
- Bäckman, L., & Karlsson, T. (1985). The relation between level of general knowledge and feeling-of-knowing: An adult age study. *Scandinavian Journal of Psychology*, 26, 249–258.
- Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, 24, 654–663.
- Cohen, G., & Faulkner, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology & Aging*, 4, 10–17.
- Coxon, P., & Valentine, T. (1997). The effects of the age of eyewitnesses on the accuracy and suggestibility of their testimony. *Applied Cognitive Psychology*, 11(5), 415–430.
- Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London B*, 302, 341–359.
- Craik, F. I., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 474–479.
- Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: Why older adults are unreliable eyewitnesses. *Psychonomic Bulletin & Review*, 13(5), 770–775.
- Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging*, 22(1), 122–133.
- Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 813–825.
- Goldsmith, M., & Koriat, A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin & B. Ross (Eds.), *Psychology of learning and motivation: Memory use as skilled cognition* (Vol. 48, pp. 1–60). San Diego, CA: Elsevier.
- Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging*, 4, 106–112.
- Holliday, R. E., Humphries, J. E., Milne, R., Memon, A., Houlter, L., & Lyons, A., et al. (2011). Reducing misinformation effects in older adults with cognitive interview. Mnemonics. *Psychology and Aging*. doi:10.1037/a0022031.
- Ikier, S., Yang, L., & Hasher, L. (2008). Implicit interference, age, and automatic versus controlled retrieval strategy. *Psychological Science*, 19, 456–461.
- Jacoby, L. L., & Hollingshead, A. (1990). Reading student essays may be hazardous to your spelling: Effects of reading incorrectly and correctly spelled words. *Canadian Journal of Psychology*, 44, 345–358.
- 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.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28.
- Karayanni, I., & Gardiner, J. M. (2003). Transferring voice effects in recognition memory from remembering to knowing. *Memory & Cognition*, 31, 1052–1059.
- Karpel, M. E., Hoyer, W. J., & Togli, M. P. (2001). Accuracy and qualities of real and suggested memories: Nonspecific age differences. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, 56, 103–110.
- Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language*, 48(4), 704–721.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103(3), 490–517.
- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, 135, 36–69.
- Lachman, J. L., Lachman, R., & Thronesbery, C. (1979). Metamemory through the adult life span. *Developmental Psychology*, 15(5), 543–551.
- Lane, S. M., Roussel, C. C., Starns, J. J., Villa, D., & Alonzo, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory*, 16(8), 836–851.
- Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4(1), 19–31.
- Lovell, E. A., & Marsh, G. R. (1985). Prediction and evaluation of memory performance by young and old adults. *Journal of Gerontology*, 40, 192–197.
- Marquié, J. C., & Huet, N. (2000). Age differences in feeling-of-knowing and confidence judgments as a function of knowledge domain. *Psychology and Aging*, 15, 451–461.
- Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology*, 17(1), 107–119.
- Naveh-Benjamin, M., & Craik, F. I. M. (1995). Memory for context and its use in item memory: Comparisons of younger and older persons. *Psychology and Aging*, 10, 284–293.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–140). New York: Academic Press.
- Pansky, A., Goldsmith, M., Koriat, A., & Pearlman-Avni, S. (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology*, 21, 303–329.
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, 15, 105–130.
- Perfect, T. J., & Stollery, B. T. (1993). Memory and metamemory performance in older adults: One deficit or two? *Quarterly Journal of Experimental Psychology*, 46A, 119–135.
- Perrotin, A., Isingrini, M., Souchay, C., Clarys, D., & Taconnat, L. (2006). Episodic feeling-of-knowing accuracy and cued recall in the elderly: Evidence for double dissociation involving executive function and processing speed. *Acta Psychologica*, 122, 58–73.
- Reder, L. M., Wible, C., & Martin, J. (1986). Differential memory changes with age: Exact retrieval versus plausible inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 72–81.
- Rhodes, M. G., & Kelley, C. M. (2005). Executive processes, memory accuracy, and memory monitoring: An aging and individual differences analysis. *Journal of Memory and Language*, 52, 578–594.

- Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence–accuracy correlation. *Journal of Applied Psychology, 81*, 587–594.
- Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process vs. content in eyewitness memory monitoring. *Journal of Experimental Psychology: Applied, 6*(3), 207–221.
- Roediger, H. L., & Gynn, M. J. (1996). Retrieval processes. In E. L. Bjork & R. A. Bjork (Eds.), *Human memory* (pp. 197–236). San Diego: Academic Press.
- Roediger, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language. Special Issue: Illusions of Memory, 35*(2), 300–318.
- Salthouse, T. A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology: Psychological Sciences, 29*–36.
- Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology, 14*(2), 299–309.
- Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling of knowing and recollection. *Consciousness & Cognition, 16*(4), 769–784.
- Thomas, A. K., & Bulevich, J. B. (2006). Effective cue utilization reduces memory errors in older adults. *Psychology and Aging, 21*(2), 379–389.
- Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning – Implications for retrieval enhanced suggestibility. *Journal of Memory and Language, 63*, 149–157.
- Thomas, A. K., Bulevich, J. B., & Dubois, S. (2011). The role of contextual information in episodic feeling of knowing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 96–108.
- Thomas, A. K., Dubois, S. J., & Bulevich, J. B. (in preparation). The dynamic interplay between the quality and quantity of contextual information in the feeling of knowing.
- Weingardt, K. R., Leonesio, R. J., & Loftus, E. F. (1994). Knowing about knowing. In J. Metcalfe & A. Shimamura (Eds.), *Viewing eyewitness research from a metacognitive perspective* (pp. 157–184). Cambridge, MA: MIT Press, Bradford Books.
- Weingardt, K. R., Toland, H. K., & Loftus, E. F. (1994). Reports of suggested memories: Do people truly believe them? In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 3–26). New York, NY, US: Cambridge University Press.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of Remember/Know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.
- Yonelinas, A. P. (2001). Consciousness, control, and confidence. The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130*, 361–379.