

# Retrieval monitoring is influenced by information value: The interplay between importance and confidence on false memory☆



Ian M. McDonough<sup>a,\*</sup>, Dung C. Bui<sup>b</sup>, Michael C. Friedman<sup>c</sup>, Alan D. Castel<sup>d</sup>

<sup>a</sup> Department of Psychology, University of Alabama, 505 Hackberry Lane, Tuscaloosa, AL 35487, USA

<sup>b</sup> Department of Psychology, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA

<sup>c</sup> Harvard Initiative for Learning and Teaching, Harvard University, 125 Mt. Auburn Street, Cambridge, MA 02138, USA

<sup>d</sup> Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095, USA

## ARTICLE INFO

### Article history:

Received 20 December 2014

Received in revised form 28 May 2015

Accepted 29 July 2015

Available online xxxx

### Keywords:

False memory

Heuristics

Memory

Metamemory

Word recognition

Value

## ABSTRACT

The perceived value of information can influence one's motivation to successfully remember that information. This study investigated how information value can affect memory search and evaluation processes (i.e., retrieval monitoring). In Experiment 1, participants studied unrelated words associated with low, medium, or high values. Subsequent memory tests required participants to selectively monitor retrieval for different values. False memory effects were smaller when searching memory for high-value than low-value words, suggesting that people more effectively monitored more important information. In Experiment 2, participants studied semantically-related words, and the need for retrieval monitoring was reduced at test by using inclusion instructions (i.e., endorsement of any word related to the studied words) compared with standard instructions. Inclusion instructions led to increases in false recognition for low-value, but not for high-value words, suggesting that under standard-instruction conditions retrieval monitoring was less likely to occur for important information. Experiment 3 showed that words retrieved with lower confidence were associated with more effective retrieval monitoring, suggesting that the quality of the retrieved memory influenced the degree and effectiveness of monitoring processes. Ironically, unless encouraged to do so, people were less likely to carefully monitor important information, even though people want to remember important memories most accurately.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Every day people encounter more information than they can possibly remember. As technology advances, even more information is available through a variety of media outlets. This increased exposure to information heightens the need for people to selectively attend to and prioritize information based on their personal goals and motivations that are most important so that they will be more likely to later remember that information. Indeed, people are quite adept at later remembering information deemed to be important (e.g., Ariel, Dunlosky, & Bailey, 2009; Castel, Benjamin, Craik, & Watkins, 2002; Loftus & Wickens, 1970). Furthermore, people also expect to remember more important information than less important information (e.g., Festini, Hartley, Tauber, & Rhodes, 2013; Friedman & Castel, 2011; Kassam, Gilbert, Swencionis, & Wilson, 2009). While information deemed important might affect how people strategically attend to or encode information, whether the importance of information affects the strategies by which people search and evaluate their memories at retrieval (i.e., retrieval monitoring) is virtually unknown.

Does event importance affect retrieval-monitoring processes? If so, what stages of retrieval monitoring are most affected? To address this question, the present study manipulated the perceived importance of information and the degree of retrieval monitoring.

Theories pertaining to retrieval monitoring propose that memorial expectations serve a key role in how people search memory and evaluate retrieved information (e.g., Gallo, 2010; Johnson, Hashtroudi, & Lindsay, 1993). For instance, people expect that visually distinctive events (i.e., events that contain many item-specific details) such as pictures will be better remembered than visually impoverished information such as words (e.g., Schacter & Wiseman, 2006). However, when the expected item-specific details cannot be retrieved for a particular event, people quickly and accurately reject the event as having occurred (e.g., "This item probably wasn't presented as a picture, because I'd remember more specific details."). In other words, people base their memory decisions on the expected qualitative characteristics of the to-be-recalled information and can avoid false memories if the retrieved information does not match those expectations. This type of retrieval-monitoring process has been referred to as a distinctiveness heuristic (Schacter, Israel, & Racine, 1999) and has been extended to conceptually distinctive relative to conceptually non-distinctive events as well (e.g., Gallo, Meadow, Johnson, & Foster, 2008; McDonough & Gallo, 2008). Thus, this reduced susceptibility to false memories following the

☆ Aspects of these data were presented at the 53rd Annual Meeting of the Psychonomic Society (2013, Toronto, Canada).

\* Corresponding author at: 505 Hackberry Lane, Tuscaloosa, AL 35487, USA.  
E-mail address: [ian.mcdonough@gmail.com](mailto:ian.mcdonough@gmail.com) (I.M. McDonough).

encoding of distinctive events leads to increased monitoring effectiveness or the degree that retrieval monitoring leads to more accurate memory decisions.

Does important information leave more distinctive memory traces (i.e., contain more item-specific details) than less important information? If so, encoding important information might lead to fewer false memories when retrieved at a later point. Investigating how the importance of information affects memory has been approached using the value-directed-remembering (VDR) paradigm (Castel, 2007; Castel et al., 2002) in which participants study a list of words paired with a number or “value” (e.g., skate 7, cheek 12, fence 3) to remember for a later memory test. Prior to viewing the lists, participants are told that they will be awarded the point value associated with each remembered word, thus making words with higher numbers more valuable to remember than words with lower numbers (in the example above, remembering “cheek” would be more valuable than remembering “skate” or “fence”). VDR studies have demonstrated that participants recall or recognize more words that were paired with higher values than words with lower values (e.g., Ariel et al., 2009; Castel, Farb, & Craik, 2007; Castel et al., 2002; Loftus & Wickens, 1970). In addition, recent unpublished research indicates that high-value items are associated with more “remember” responses than low-value items, consistent with this idea (Cohen, Rissman, Harbert, Castel, & Knowlton, 2013; Hennessee, Cohen, Castel, & Knowlton, 2014).

While better memory and more “remember” responses for high-value items relative to low-value items is consistent with a distinctiveness account, these findings are not sufficient to argue that high-value items will lead to more effective retrieval monitoring than low-value items. For example, McDonough and Gallo (2008) showed that associating a presented object with a personal autobiographical memory resulted in both greater recognition memory and more frequent “remember” responses than judging whether an object was made in a factory, ostensibly due to the greater degree of elaboration during encoding. Consistent with retrieval-monitoring theories, searching memory for the autobiographical items was associated with fewer false memories than searching memory for factory items. Critically, repeating the factory judgments multiple times equated recognition memory and “remember” responses, but false memories continued to be lower when participants searched for the more distinctive types of events (i.e., autobiographical items). Using a different approach, Scimeca, McDonough, and Gallo (2011) showed that repeatedly presenting words led to greater recognition memory and more frequent “remember” responses than not repeating words, but false memories were *not* reduced when searching memory for repeated words compared with non-repeated words. They further showed that subjective ratings of item distinctiveness did not differ between repeated and non-repeated words. Together, these studies suggest that relative differences in memory strength and “remember” responses do not predict when retrieval monitoring will be effective, as measured through a reduction in false memories.

An alternative possibility is that important events lead people to try harder to remember the events by engaging in retrieval monitoring to a greater extent than less important events. That is, people might initiate multiple search attempts and take time to carefully evaluate the retrieved details, but nevertheless continue to be susceptible to false memories. This idea of monitoring engagement is orthogonal to the idea of monitoring effectiveness. For instance, people often take more time to respond when searching memory for non-distinctive events (i.e., more engagement), but the outcome of this additional effort often is in vain because memory accuracy is still poor (i.e., poor effectiveness; e.g., Gallo, Kensinger, & Schacter, 2006; Gallo, McDonough, & Scimeca, 2010).

Some experimental evidence is consistent with the idea that retrieval monitoring may not be affected by the importance of information. A study by Kassam et al. (2009) showed that people expected that knowing the importance of information both before and after encoding should impact subsequent memory, but they found that only knowing information

before encoding impacted subsequent memory. This finding suggests that important information has large effects on attention and encoding strategies, but little or no effect on retrieval strategies. In addition, a recent study showed that processing information associated with more important information had the unintended consequence of creating more false memories relative to processing less important information when participants studied lists of semantic associates, known to create high levels of false memories (Bui, Friedman, McDonough, & Castel, 2013). While neither study isolated potential contributions of retrieval monitoring as a function of importance, these initial sources of evidence question the degree to which retrieval-monitoring processes effectively reduced susceptibility to false memories following the encoding of important information. In three experiments, we use well-established methods known to engage and manipulate retrieval-monitoring processes to reveal how those processes are affected by encoding information associated with different values.

### 1.1. Experiment 1

The primary aim of Experiment 1 was to test the degree to which encoding varying levels of importance affects retrieval monitoring. We used the criterial recollection task (Gallo, 2013; Gallo, Weiss, & Schacter, 2004)—a variant of traditional source memory tests—in which participants study different categories of stimuli at encoding (e.g., pictures and words), and then at test, they are oriented toward a specific category of stimuli (e.g., “Was this cue presented previously as a picture?”). To correctly respond, participants must recollect item-specific details for the queried format similar to source memory tests. In this way, criterial-recollection tests encourage retrieval monitoring of item-specific details across all items (e.g., Lindsay & Johnson, 1989; Multhaup, De Leonardis, & Johnson, 1999). However, unlike traditional source memory tests, criterial-recollection tests allow for the assessment of how effective these processes are during retrieval, as measured by the relative level of source misattributions across the different tests. That is, by having people assess each source separately, people can adjust their memorial expectations and subsequent retrieval-monitoring processes. This task has revealed different degrees of retrieval-monitoring effectiveness between many different types of stimuli (for review, see Gallo, 2013). Notably, this task uses unrelated words to minimize the effect of associative activation or the formation of gist traces (Brainerd & Reyna, 1998; Deese, 1959; Roediger & McDermott, 1995; Schacter, Verfaellie, & Pradere, 1996), and encourages retrieval monitoring (on average) for each category or stimulus type.

In Experiment 1, participants first studied a block of words (not paired with a value) to serve as lures during the criterial recollection tests (for a similar method, see Scimeca, McDonough, & Gallo, 2011).<sup>1</sup> Following this familiarization phase, participants studied words paired with low, medium, and high values (the value-study phase). Because targets (paired with values) and lures (not paired with values) were each only presented once, they should be relatively matched on familiarity. Participants then received three criterial-recollection tests: a low-value test, a medium-value test, and a high-value test. On each test, participants were oriented to one of the values and were asked whether the presented word was presented with the criterial value

<sup>1</sup> A modified version of the original criterial recollection task was used in Experiment 1. The original criterial recollection task has participants discriminate between the same two sources, but the class of items being searched for is different on each test. For example, cues for previously studied words and pictures would be presented on each test, but participants would either be asked if the cue was previously seen as a word on one test or as a picture on the other test. Thus, words studied in non-target values would serve as lures (e.g., words on the picture test). However, this approach leads to differential memory strength for lures across the different tests (i.e., picture lures on a word test would be more likely to be retrieved than word lures on a picture test). Instead, we adopted an approach used and validated by Scimeca et al. (2011) that implemented a familiarization phase first, which allowed lures to be identical across the three value tests. This modified version is conceptually similar to the original criterial recollection tests, but is more tightly controlled.

(i.e., high-value words) or not (i.e., words from the familiarization phase with no value or words that were not previously presented). Participants should respond “yes” if they recollect the value category of the word. Based on a distinctiveness account of value, participants should make fewer source misattributions on the high-value test than the low-value test due to more effective retrieval monitoring on the high-than low-value test. However, to the extent that encoding important information does not lead to more effective monitoring, perhaps because it simply leads to more effort or engagement during retrieval monitoring, source misattributions may not differ across value tests. Furthermore, we focused primarily on the effects of false memories both because the criterial recollection task is most sensitive to false memory effects and because the effects of value on true memories are sometimes reduced on recognition memory tests (Spaniol, Schain, & Bowen, 2014).

## 2. Method

### 2.1. Participants

One hundred and twenty seven participants (89 females and 38 males; mean age = 34.6 years, ages ranged from 18–72) were recruited from Amazon’s Mechanical Turk (for a recent overview, see Mason & Suri, 2012) to take part in this study for monetary compensation. All participants reported proficiency in English and resided in the United States.

### 2.2. Materials and design

Stimuli were 192 common and unrelated object words from Gallo et al. (2004) and were divided into 12 lists of 16 words. Lists of studied items were presented in two phases (Fig. 1). The familiarization phase consisted of six lists of 96 words that were presented in a different random order for each participant and these words were not associated with any value. Half of these words (three lists) were only studied in the familiarization block to serve as familiar lures for the subsequent memory tests (i.e., words that were studied, but not with any of the criterial values). The other half of the words were studied both in the familiarization phase and in a second phase (the value-study phase) in which the words were paired with point values. Specifically, of the three additional lists, one list was paired with low values (1–32), one with medium values (33–64), and one with high values (65–96). These “both items” were included so that participants could not reject words associated with no value at test by recollecting that the item had been presented in the familiarization phase (i.e., a recall-to-reject strategy). In this second phase, an additional three lists of words that were not studied in the familiarization phase also were paired with low, medium, or high values (i.e., criterial targets). Thus, a total of nine lists of words were studied and classified as either familiar lures, “both” targets, or criterial

targets. The last three lists were not studied and were classified as new lures for the subsequent memory tests.

Participants completed three different memory tests (low-value test, medium-value test, and high-value test) with the order of each test counterbalanced across participants. During each test, participants were presented with 64 words in a randomized order. There were 16 words of each type in each test (“both” targets, criterial targets, familiar lures, and new lures). For example, the high-value test contained words that were studied with high values (criterial targets), no value (familiar lures), both high values and no values (“both” targets), or were not studied at all (new lures). Words were only tested once, and a prompt was presented at the top of the screen on each test (e.g., “High value 65–96?”) to remind participants of the values encompassed by the current test. Each set of words was counterbalanced across the study and test conditions.

### 2.3. Procedure

Participants read all instructions on the computer screen and were informed that they would be studying a list of words (the familiarization phase). Following the familiarization phase, participants were informed that they would be studying additional words that were paired with a number ranging from 1 to 96 (e.g., “compass 12”), and that the point values indicated how important each word was to be remembered. Participants were further told that their goal was to maximize their “memory score” by learning and remembering as many high-value words as possible, but that medium-value and low-value words would also contribute to their score (see also Castel et al., 2002). Participants also were informed that each word would appear on the screen for 3 s and after all words had been presented, they would be presented with a new list of words and decide whether or not each word was previously displayed on the screen. Participants were told not to use any external aids to help remember the words.

After the value-study phase, participants received three criterial-recollection tests, with the order of the tests counterbalanced across participants. On the high-value test, participants were told that they would be tested on words that had either been studied as high-value words only, as words with no value, as both high-value and no-value words, or were not studied. Participants were instructed to press “yes” if they remembered seeing the word paired with a high value (i.e., it might have been studied as a high-value word only, or both as a high-value word and not paired with a value), or otherwise press “no.” Participants were told that remembering whether or not they had studied the word without a value was irrelevant on the test, because some words were studied with and without a value, and others were not. The instructions for the other tests were similar, except they were adjusted to correspond to medium and low values rather than high values. Demographic information was collected last.

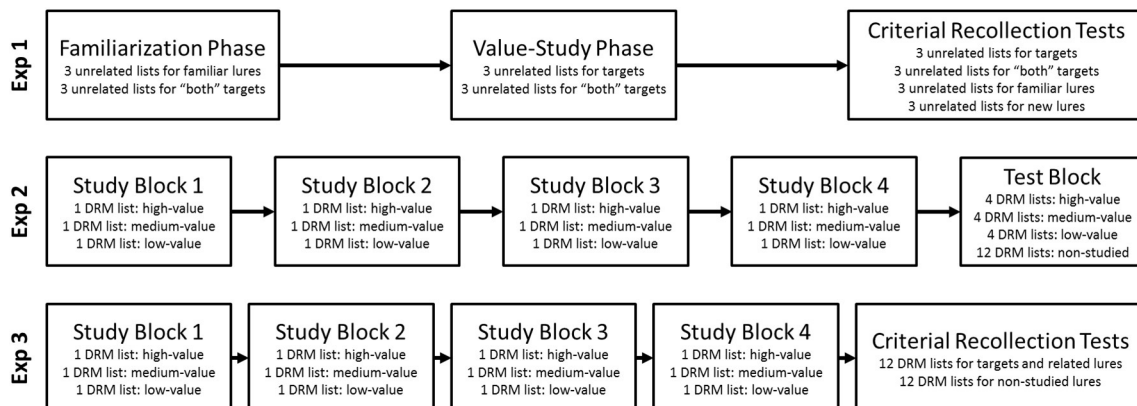


Fig. 1. Schematic of experimental design for Experiment 1, Experiment 2, and Experiment 3. Exp = Experiment.



### 3. Results and discussion

Because we had a wide age range (18–72 years,  $M = 34.59$ ), reported means and analyses were conducted on values adjusted for age by using the residualized values after regressing age on our dependent variables of interest. Three participants were excluded because response times for over two-thirds of the test words were faster than 300 ms and three additional participants were excluded because their mean hit rate for criterial targets was 10% or less, suggesting that they may have not been engaging in the task (final  $N = 121$ ).

Correct source attributions to targets and incorrect source misattributions to lures are shown in Fig. 2A and B, respectively. A 4 (item type: “both” targets, criterial targets, familiar lures, new lures)  $\times$  3 (value: low, medium, high) repeated measures ANOVA revealed a main effect of item type (“both” targets > criterial targets > familiar lures > new lures),  $F(3, 360) = 190.74$ ,  $MSE = .047$ ,  $p < .001$ ,  $\eta_p^2 = .61$ , which was qualified by an item type  $\times$  value interaction,  $F(6, 720) = 8.65$ ,  $MSE = .017$ ,  $p < .001$ ,  $\eta_p^2 = .07$ . The interaction indicated that while the proportion of correct source attributions to “both” targets and criterial targets increased with value, the source misattributions to familiar and new lures decreased with value. Follow-up  $t$ -tests confirmed that for “both” targets, correct source attributions were greater for words on the high-value than low-value test ( $t(120) = 2.08$ ,  $SEM = .019$ ,  $p = .04$ ,  $d = .20$ ) and for criterial targets, correct source attributions were greater on the high-value than medium-value test and low-value test ( $t(120) = 2.76$ ,  $SEM = .020$ ,  $p = .007$ ,  $d = .25$  and  $t(120) = 3.04$ ,  $SEM = .019$ ,  $p = .003$ ,  $d = .29$ , respectively). For familiar lures, source misattributions were lower on the high-value than medium-value test and low-value test ( $t(120) = 3.61$ ,  $SEM = .019$ ,  $p < .001$ ,  $d = .33$  and  $t(120) = 2.16$ ,  $SEM = .020$ ,  $p = .03$ ,  $d = .20$ , respectively) and for new lures, source misattributions were lower on the high-value than medium-value test and low-value test ( $t(120) = 3.60$ ,  $SEM = .017$ ,  $p < .001$ ,  $d = .30$  and  $t(120) = 2.70$ ,  $SEM = .018$ ,  $p = .008$ ,  $d = .23$ , respectively).

As predicted by the distinctiveness account of retrieval-monitoring theories, when encouraged to monitor words across all values, participants more effectively used retrieval-monitoring processes for high-value than low-value words akin to a distinctiveness heuristic (e.g., Schacter et al., 1999). This conclusion stems from the finding that people had fewer source misattributions on the high-value test relative to the medium-value and low-value test.

#### 3.1. Experiment 2

The findings from Experiment 1 suggest that when people search memory for information they deem important, they demand more item-specific or distinctive recollections to endorse an event as having occurred. Having established this basic result, we turn our focus now to a recent study that appeared to conflict with these results. Specifically, Bui et al. (2013) showed that semantically-related words (i.e., DRM lists; Deese, 1959; Roediger & McDermott, 1995) associated with higher values were accompanied by greater false recall compared with semantically-related words associated with lower values. One explanation for their findings is that encoding higher values engendered more relational processing (elaborating on features common to studied items; Hunt & Einstein, 1981; Hunt & McDaniel, 1993), which in turn increased false memories via spreading activation between preexisting conceptual representations in participants' mental lexicon (e.g., Roediger, Balota, & Watson, 2001) or via the construction of a gist representation (e.g., Brainerd & Reyna, 1998; Schacter et al., 1996). However, as shown in Experiment 1, encoding value affected not only encoding processes, but also retrieval processes. Thus, an alternative explanation for the findings from Bui et al. (2013) could be that participants more effectively monitored items at retrieval related to low value than high value in that particular context (i.e., after studying semantically-related words). This alternative explanation, however, is inconsistent with the

findings in Experiment 1; participants more effectively utilized retrieval monitoring processes for items related to high value than low value.

There are at least two major factors that could lead to different effects of value on false memories. The first factor is the semantic relatedness of the stimulus materials, which will be addressed more closely in Experiment 3. The second factor is the type of memory test, which is the focus of Experiment 2. Depending on the type of memory test, participants might be more or less inclined to carefully inspect memory for item-specific details that could lead to effective rejection of false memories. Previous studies suggest that requiring participants to inspect their memories more carefully (e.g., using source memory tests compared with “old/new” recognition tests) leads to decisions that rely less on familiarity and more on item-specific details (e.g., Dodson & Johnson, 1993; Jacoby, Kelley, Brown, & Jasechko, 1989; Jacoby, Woloshyn, & Kelley, 1989; Smith, Hunt, & Gallagher, 2008). In the criterial recollection tests used in Experiment 1, participants essentially search memory for source information (e.g., high value), thereby encouraging participants to carefully inspect memory for item-specific details diagnostic of the occurrence of the event, rather than mere familiarity or gist that the item had been previously seen (e.g., Gallo, 2013). In contrast, Bui et al. (2013) used a free recall paradigm in which any item that seemed “old” was likely reported. In this case, information related to the gist of the event that may not be associated with a large degree of item-specific details (e.g., the point value it was associated with) would likely be recalled (e.g., Reyna & Brainerd, 1995; Roediger & McDermott, 1995). Because false memories associated with high-value lists likely were more familiar and/or had greater gist strength, infrequent retrieval monitoring would lead to more frequent false memories as value increased.

Also related to these ideas is research showing that people are less likely to engage in retrieval-monitoring processes for information recalled vividly or recalled with high confidence. Jacoby and Hollingshead (1990) suggested that during recall, people search memory and generate a potential response, then engage in recognition (and monitoring) processes to judge whether that item had previously been studied. However, they found that for items that were recalled fluently, additional recognition (and monitoring) processes were less likely to be engaged. Thus, one possibility is that false memories associated with high-value lists were retrieved more vividly (i.e., false recollections) or with higher confidence than false memories associated with low-value lists. Under these conditions, retrieval monitoring might actually be more likely to occur for low-value items because those items should be least vividly recalled and/or recalled with less confidence, thus creating doubt that the retrieved information represented a true memory (e.g., Asp & Tranel, 2012). Indeed, responses accompanied by lower confidence judgments often are associated with stricter retrieval criteria (i.e., more “new” responses; e.g., Rotello & Macmillan, 2007), thus reducing false memories.

The primary goals of Experiment 2 were 1) to conceptually replicate the effects of value on false memory found in Bui et al. (2013) using a recognition test and 2) to manipulate the degree to which participants would engage retrieval-monitoring processes by altering the test instructions (i.e., with inclusion instructions). We used recognition memory tests rather than free recall (as in Bui et al., 2013) to allow for better comparison with Experiment 1 as well as to generalize the value-false memory effects previously found. Experiment 2 used lists of semantic associates (i.e., DRM lists) with one group of participants receiving standard “old/new” recognition instructions and another group receiving inclusion recognition instructions that should minimize or reduce the engagement of retrieval-monitoring processes (e.g., Brainerd & Reyna, 1998; Hege & Dodson, 2004; Hunt, Smith, & Dunlap, 2011; Pierce, Gallo, Weiss, & Schacter, 2005). Specifically, the group receiving inclusion instructions were instructed to endorse words as “old” if each word was related to the previously studied words, regardless of whether the item was actually studied. These inclusion instructions differed from standard instructions, in which participants should endorse words as “old” only if the word was previously studied. Inclusion instructions required remembering the themes of the studied lists rather than item-

specific details, and thus retrieval-monitoring processes should have been less necessary. Thus, the difference in false recognition between the two instruction conditions should reveal when monitoring processes were engaged (i.e., for low-value trials, high-value trials, or all trials).

Based on the different likelihoods of engaging in retrieval-monitoring processes between “old/new” recognition tests and source memory tests (or in this case criterial recollection tests), we predicted the opposite direction of results as in Experiment 1. Overall, critical lures associated with higher-value lists should have greater levels of false recognition because associative activation/gist processing would be greatest in high-value than low-value lists, thus conceptually replicating Bui et al. (2013). However, the critical question was whether the difference in false recognition between the two instruction conditions would be greatest for high-value items or low-value items. To the extent that vividness of details and/or confidence affects when people engage in retrieval-monitoring for item-specific details, larger differences should be found for low-value items because these items should engender the least vivid details and/or confidence, thus encouraging the engagement of retrieval-monitoring processes. Alternatively, participants better remember the value of items encoded with high-value than low-value (Castel et al., 2007), and therefore might be more likely to engage retrieval monitoring selectively for high-value items. In other words, people might be more motivated to accurately remember high-value items, thus leading to larger false recognition differences for high-value items.

## 4. Method

### 4.1. Participants and design

Ninety-six participants (60 females and 36 males; mean age = 35.9 years, ages ranged from 18–66) were recruited from Amazon’s Mechanical Turk to take part in this study for monetary compensation. All participants reported proficiency in English and resided in the United States. Test instructions (standard, inclusion) were manipulated between subjects, and value (low, medium, high) was manipulated within subjects.

### 4.2. Materials

Twenty-four word-lists were used in this experiment, and were the same as those used in the Roediger and McDermott study (1995, Experiment 2). Each list contained 14 words, all of which were semantically associated with a word not included on the list and, consequently, not shown to participants (i.e., the critical lure). Four word lists were associated with low values (1–14), four lists were associated with medium values (15–28), four lists were associated with high values (29–42), and the remaining 12 lists were non-studied and served as non-critical lures at test (Fig. 1). Within each list the distribution of values was randomized across the words such that each word was given its own unique value. Value assignment was counterbalanced across participants such that all lists were at one point assigned to low, medium, and high values or were used as control lists. The lists were presented in four blocks such that three lists (one of each value type) were interleaved within each block. This method was chosen to strike a balance between differential processing across the value types and successful activation or gist formation of the critical lures.

All test words were presented in a different random order for each participant and consisted of one critical lure from each studied list, two targets (i.e., the first and eighth strongest associates of the critical lure) from each studied list, one non-critical lure from each non-studied list, and two non-critical targets (i.e., the first and eighth strongest associates of the non-critical lures) from each non-studied list. Thus, a total of 24 targets, 24 non-critical targets, 12 critical lures, and 12 non-critical lures were presented at test.

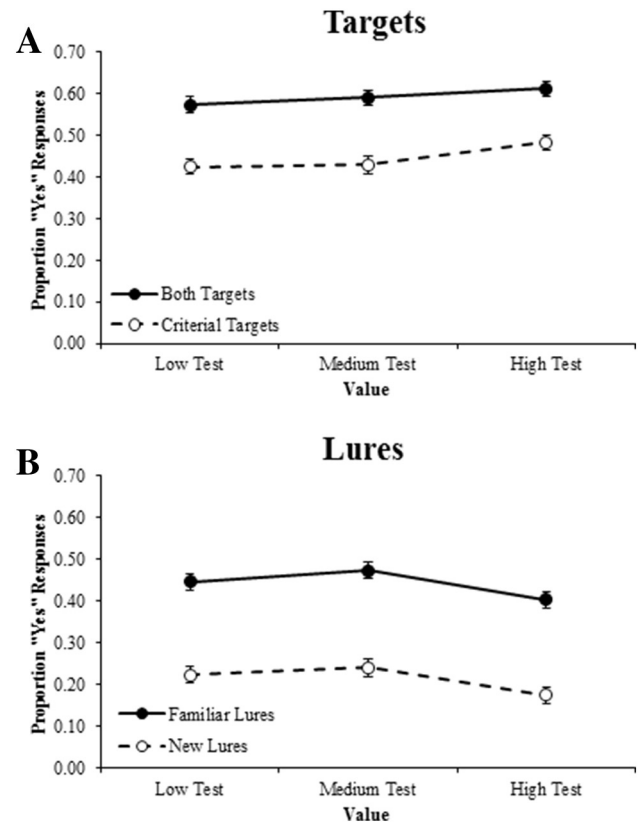


Fig. 2. Memory results for Experiment 1. Mean proportion of source attributions to targets (Panel A) and source misattributions to lures (Panel B) as a function of test (low, medium, and high value test). Error bars represent the standard error of the mean.

### 4.3. Procedure

The procedure for the study phase was similar to the value-study phase in Experiment 1, with the exception that values ranged from 1–42. After all the study lists were presented, a recognition memory test was given. Participants in the standard-instruction condition were asked whether each word was previously presented (“yes” or “no”). Participants in the inclusion-instruction condition were instructed to endorse words as “old” if they thought the word was presented (i.e., they saw the word at study), or was a related word to one that was presented. Otherwise, if they did not remember the item from any of the lists and it looked unrelated to the previously studied lists, they were instructed to respond “new.” Lastly, demographic information was collected.

## 5. Results and discussion

Two participants from the standard condition were excluded because response times for over two-thirds of the test words were faster than 300 ms (final  $N = 94$ ). As in Experiment 1, we report age-adjusted means and analyses.

The proportion of “old” responses to targets and lures are shown in Fig. 3A and B, respectively.<sup>2</sup> A 2 (item type: targets, critical lures)  $\times$  3 (value: low, medium, high)  $\times$  2 (instruction: standard, inclusion) mixed-factor ANOVA revealed a main effect of item type,  $F(1, 92) = 13.06$ ,  $MSE = .060$ ,  $p < .001$ ,  $\eta_p^2 = .12$ , such that the proportion of “old” responses were greater for targets than lures in both instruction

<sup>2</sup> Hits to targets and false alarms to critical lures were not corrected by false alarms to non-critical targets and non-critical lures in Experiment 2. We expected an inflated proportion of “old” responses to all items due to the instruction manipulation that might be masked or undermined by correcting for these values. Instead, false alarms to these non-critical items were separately analyzed as a function of instruction.

conditions. There was also a main effect of value (higher > lower),  $F(2, 184) = 8.28$ ,  $MSE = .044$ ,  $p < .001$ ,  $\eta_p^2 = .08$ , and a main effect of instruction (inclusion > standard),  $F(1, 92) = 4.02$ ,  $MSE = .19$ ,  $p = .05$ ,  $\eta_p^2 = .04$ , which were qualified by a value  $\times$  instruction interaction,  $F(2, 184) = 3.84$ ,  $MSE = .044$ ,  $p = .02$ ,  $\eta_p^2 = .04$ .

We first note that higher value was associated with a greater proportion of “old” responses for targets and lures than lower value in the standard-instruction condition, conceptually replicating Bui et al. (2013), all  $p$ 's < .05. Follow-up  $t$ -tests between instruction conditions revealed that the proportion of “old” responses for low-value targets using inclusion-instructions was greater than low-value targets using standard-instructions,  $t(92) = 2.77$ ,  $SEM = .05$ ,  $p = .007$ ,  $d = .57$ . Similarly, the proportion of “old” responses for low-value and medium-value lures using inclusion instructions was greater than for low-value and medium-value lures using standard instructions ( $t(92) = 2.34$ ,  $SEM = .06$ ,  $p = .02$ ,  $d = .48$  and  $t(92) = 2.01$ ,  $SEM = .06$ ,  $p = .05$ ,  $d = .41$ ). Differences as a function of instruction condition were not significant for high-value targets or lures (all  $p$ 's > .22).

An ANOVA was conducted on the unrelated lures (non-critical targets and non-critical lures). To the extent that retrieval monitoring was primarily engaged for words with low vividness or confidence, unrelated lures might not be affected by the manipulation; people should confidently report that these words were not previously presented (i.e., little to no memory should be associated with new words). However, to the extent that the inclusion instructions reduced retrieval monitoring across all words, we might expect to see a global increase in the proportion of “old” responses to these unrelated lures in the inclusion condition compared with the standard condition. A 2 (lure type: non-critical target, non-critical lure)  $\times$  2 (instruction: standard, inclusion) ANOVA revealed a main effect of lure type,  $F(1, 92) = 12.52$ ,  $MSE = .013$ ,  $p = .001$ ,  $\eta_p^2 = .12$ , such that the proportion of “old” responses was greater for non-critical lures ( $M = .29$ ,  $SE = .03$ ) than non-critical targets ( $M = .24$ ,  $SE = .02$ ), owing to the fact that non-critical lures have a higher baseline frequency than non-critical targets. No significant effect of instruction was found, ( $p = .09$ ), suggesting that the test instruction manipulation selectively affected retrieval monitoring for targets and critical lures.

Overall, participants did not exhibit an increase in “old” responses for high-value items with inclusion instructions, suggesting that minimal retrieval monitoring was occurring for high-value items under standard conditions (but note that people did engage in retrieval monitoring for these items when encouraged to do so, as in Experiment 1). These results are inconsistent with the idea that retrieval monitoring is always more effective for events encoded with high-value compared with low-value. Rather, they are consistent with the idea that participants remembered the theme of the low-value lists, but the retrieval of less vivid or less confident memories (i.e., familiar or gist memory) resulted in the engagement of retrieval-monitoring processes based on item-specific details selectively for low-value information. Thus, Experiment 2 provides preliminary evidence that retrieval-monitoring processes are less likely to be invoked for high-value information in the context of an “old/new” recognition test, potentially explaining why high-value items were associated with greater false memories in the DRM paradigm as reported by Bui et al. (2013).

### 5.1. Experiment 3

Thus far, Experiment 1 suggests that when people search memory for item-specific details (i.e., in the criterial recollection task), higher value is associated with more effective retrieval monitoring as evidenced by reduced source misattributions. However, the criterial recollection task encourages participants to engage in retrieval monitoring for all classes of items. In contrast, “old/new” recognition tasks do not encourage retrieval monitoring for high-value items to the same extent as revealed by the inclusion instructions in Experiment 2. Rather, retrieval-monitoring processes might be engaged on a trial-by-trial

basis depending on the level of vividness or confidence one has in the memory. That is, less vivid or confident memories might encourage more retrieval monitoring, whereas more vivid or confident memories might encourage less retrieval monitoring (e.g., Asp & Tranel, 2012; Jacoby & Hollingshead, 1990; Rotello & Macmillan, 2007).

While the results of Experiment 2 are consistent with this interpretation, the use of semantically-related words in the value-directed paradigm introduces alternative hypotheses. Because participants are instructed to earn the most points by remembering high-value items at the expense of other items, they might be more motivated to guess that semantically-related lures associated with high-value lists were previously seen (i.e., endorsed as “old”) than for low-value lists. This alternative explanation is still in line with the idea that participants are less likely to engage in extensive retrieval monitoring for high-value items compared with low-value items, but the reason would not be due to high vividness/confidence of the memory, but rather to a guessing strategy to earn the most points regardless of confidence. For example, if participants were presented with the word “sweet” at test they might have had no recollection of or familiarity with that word, but they might have remembered that a related word (e.g., “candy”) was presented and, because there is no penalty for guessing, they might have been more likely to claim that item was “old.” Note that if one takes this example to an extreme, a person may simply endorse all items of any value on a recognition test to earn the most points. Clearly, this latter scenario was not occurring; participants were basing their responses on memory signals, but likely were trying to strike a balance between earning the most points and guessing. Nevertheless, a question remains as to whether vividness or confidence in the memory influences the degree to which people engage in extensive retrieval monitoring or whether reward-induced response strategies are influencing the extent of retrieval monitoring. The first goal of Experiment 3 was to gather confidence judgments on each trial to test whether lures associated with high-value lists were endorsed with higher confidence compared with lures associated with lower-value lists.<sup>3</sup>

The second (and related) goal of Experiment 3 was to test whether confidence influenced retrieval monitoring on a trial-by-trial basis. While most studies have investigated aggregate differences as a function of condition, this metric assumes that individuals are engaging in the same types of processes for each trial (at least on average). However, there also is substantial variability in the efficiency and effectiveness of monitoring processes within a person (i.e., intra-individual differences). Hierarchical linear models (HLMs) allow for the assessment of trial-by-trial differences when explaining the processes of interest. Specifically, each item on a memory test may be retrieved with different levels of vividness and confidence, and we predicted that this trial-by-trial variability would affect the degree of retrieval monitoring engaged in, in addition to average (aggregate) effects seen by value type. Thus, within each test, we predicted that lower confidence would be associated with more extensive retrieval monitoring, which would reveal the effects of value on source misattributions (i.e., lower misattributions on the high-value test than low-value test).

To test these two goals, we used a hybrid design such that DRM lists were presented at encoding (thus using semantically-related words) as in Experiment 2, but people were asked to retrieve details pertaining to specific sources of information analogous to criterial recollection tests as in Experiment 1. This design allowed us to investigate the combination of encoding semantically-related words and encouraging retrieval-monitoring across all value types. We predicted that for unrelated lures, we would see a similar reduction in source misattributions on the high-value test compared with the low-value test (as in Experiment

<sup>3</sup> We chose to gather ratings of confidence rather than vividness of each memory because our hypotheses were most congruent with the idea that confidence of the total memory output affects the engagement of retrieval-monitoring processes. Vividness is one of multiple factors that influence confidence (e.g., Busey, Tunnicliff, Loftus, & Loftus, 2000; Reinitz, Peria, Séguin, & Loftus, 2011), and so seemed most appropriate.



1), thus showing evidence of more effective retrieval when retrieval monitoring is encouraged across all value types (on average). For related words, we predicted a muted effect of value more consistent with Experiment 2; higher value should be associated with greater source misattributions, but due to the opposition of retrieval-monitoring processes (which are more likely to occur, on average on these memory tests), these effects should be considerably less strong than in Experiment 2.

## 6. Method

### 6.1. Participants

One hundred and one participants (76 females and 25 males; mean age = 36.7 years, ages ranged from 18–66) were recruited from Amazon's Mechanical Turk to take part in this study for monetary compensation. All participants reported proficiency in English and resided in the United States.

### 6.2. Materials and design

The materials and study phase were the same as used in Experiment 2 (standard-instruction condition). The study phase was followed by three memory tests similar to the criterial-recollection tests in Experiment 1, with the order of each test counterbalanced across participants (Fig. 1). The test phase consisted of 24 word lists with 14 words each, four of which were paired with low values, four with medium values, four with high values, and 12 were not studied. Thus, each test included items from four of the studied lists and four from the non-studied lists. Specifically, participants were presented with 24 words in a randomized order, which consisted of the eight criterial targets (i.e., the first and eighth strongest associates of the critical lure for each list), eight non-critical targets (i.e., the first and eighth strongest associates of the non-critical lures for each list), four critical lures, and four non-critical lures (i.e., critical lures in non-presented lists). Unlike traditional criterial-recollection tests, no “both” items were included on these tests because of the different nature of the lures (i.e., semantically-related words). Words were only tested once, and a prompt was presented at the top of the screen on each test (e.g., “Was this word presented with a high value 29–42?”). Confidence ratings were made using a sliding scale with a point centered in the middle by default, which could be moved to the right to indicate greater confidence or left to indicate less confidence. Each point on the scale represented a value from 0 (0% sure) to 100 (100% sure).

### 6.3. Procedure

The study-phase procedure was the same as used in Experiment 2. After the study phase, participants received three memory tests similar to Experiment 1 with the addition of confidence judgments. Specifically, participants were instructed to indicate how confident they were in each memory judgment by using the sliding scale such that moving the point to the left would indicate less confidence and to the right would indicate greater confidence. Demographic information was collected at the very end.

### 6.4. Analysis

In addition to the standard analyses conducted in the previous experiments, memory responses also were analyzed on a trial-by-trial basis and as a function of confidence. First, all confidence responses were standardized (i.e., z-scored) within each participant. Then, item type (targets, critical lures, unrelated lures)<sup>4</sup>, value (low, medium, high), and confidence were entered into an HLM as trial-level predictors

of memory performance, with participants being coded as independent random effects. A small subset (less than 1%) of participant responses was removed from the analysis due to non-responses. Standardized confidence was coded as a continuous variable, while item type and value were coded as categorical variables. Item type was dummy coded as 1, 0, and –1 for targets, critical lures, and unrelated lures, respectively. Value was coded as 1, 0, and –1 for the high-value, medium-value, and low-value tests, respectively. The dependent variable (i.e., memory performance), due to its dichotomous nature, used a standard logit-link function.

## 7. Results and discussion

One participant was excluded because response times for over half of the test words were faster than 300 ms and two participants were excluded because mean criterial hits were less than 10% (final N = 98). As in the previous experiments, we report age-adjusted means and analyses. Mean results, regardless of confidence, are reported first to mirror the analyses in the previous experiments. We then report the value effects as a function of confidence to test the extent that confidence moderated the value effects.

### 7.1. Value effects on memory

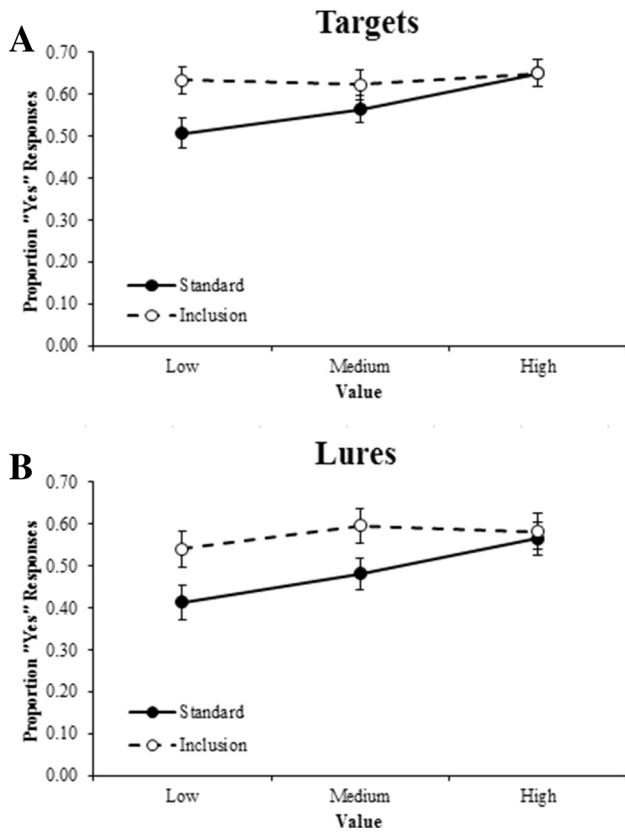
Correct and incorrect source attributions are presented as a function of item type and value in Fig. 4A. A 3 (item type: targets, related lures, unrelated lures) × 3 (value: low, medium, high) repeated measures ANOVA revealed a main effect of item type (targets > related lures > unrelated lures),  $F(2, 194) = 57.10$ ,  $MSE = .065$ ,  $p < .001$ ,  $\eta_p^2 = .37$ , which was qualified by an item type × value interaction,  $F(4, 388) = 2.79$ ,  $MSE = .033$ ,  $p = .03$ ,  $\eta_p^2 = .03$ . The interaction indicated that while correct source attributions to targets and incorrect source misattributions to related lures numerically increased with value, none of the increases were significant (all  $p$ 's > .26). In contrast, source misattributions to unrelated lures decreased significantly with value, such that misattributions on the high-value test were lower than on the medium and low-value test ( $t(97) = 2.49$ ,  $SEM = .023$ ,  $p = .01$ ,  $d = .22$  and  $t(97) = 2.40$ ,  $SEM = .028$ ,  $p = .02$ ,  $d = .25$ , respectively).

Consistent with our predictions, the value effects were not only blunted, but completely eliminated for targets and related lures, likely due to the opposing retrieval-monitoring processes on associative activation/gist processing. Critically, even when DRM lists were used, we see effects of value on the effectiveness of retrieval monitoring for unrelated lures; retrieval monitoring was more effective on the high-value test than low-value test.

### 7.2. Value effects on confidence

Mean confidence values as a function of value and item type (regardless of response) are shown in Fig. 4B. A 3 (item type: targets, related lures, unrelated lures) × 3 (value: low, medium, high) repeated measures ANOVA revealed a main effect of item type,  $F(2, 194) = 8.96$ ,  $MSE = 110.00$ ,  $p < .001$ ,  $\eta_p^2 = .09$ , such that targets were endorsed with marginally higher confidence than related lures ( $t(97) = 1.73$ ,  $SEM = .69$ ,  $p = .09$ ,  $d = .07$ , and related lures were endorsed with higher confidence than unrelated lures ( $t(97) = 2.66$ ,  $SEM = .90$ ,  $p = .009$ ,  $d = .13$ ). A main effect of value,  $F(2, 194) = 8.25$ ,  $MSE = 228.27$ ,  $p < .001$ ,  $\eta_p^2 = .08$ , indicated that words on the high-value test were endorsed with higher confidence than words on the medium- and low-value tests ( $t(97) = 2.93$ ,  $SEM = 1.29$ ,  $p = .004$ ,  $d = .20$  and  $t(97) = 3.71$ ,  $SEM = 1.30$ ,  $p < .001$ ,  $d = .25$ , respectively). The interaction was not significant ( $p = .24$ ). This analysis shows that participants remembered high-value words (targets and lures) with more confidence and is consistent with the idea that value influenced the vividness or quality of the memory, and in turn retrieval-monitoring processes, rather than simply affecting guess responses to earn more points.

<sup>4</sup> Non-critical targets and non-critical lures were averaged together for simplicity and labeled as unrelated lures.

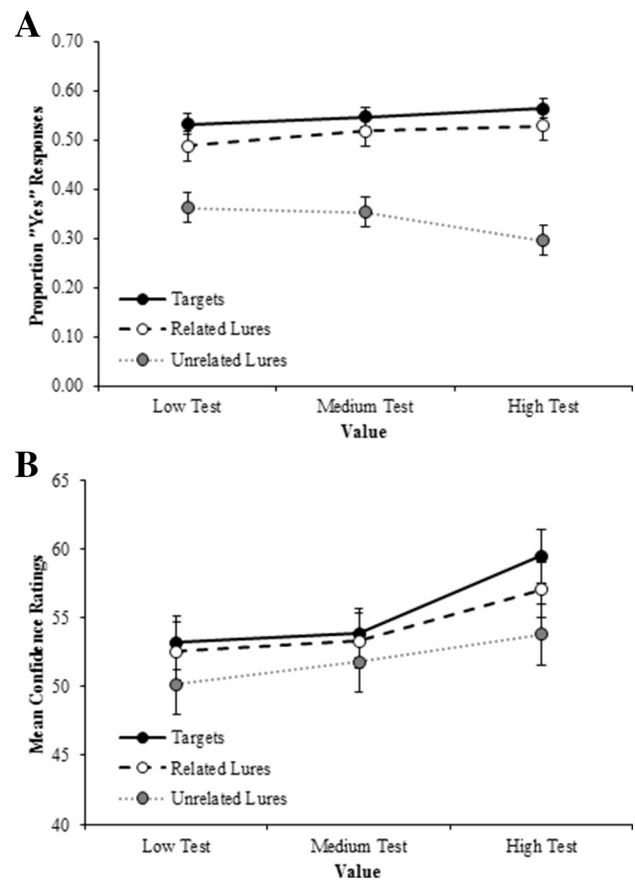


**Fig. 3.** Memory results for Experiment 2. Mean proportion of “old” responses to targets (Panel A) and lures (Panel B) for the standard instruction condition (solid line, black circles) and the inclusion-instruction condition (dashed line, white circles) as a function of binned point values. Error bars represent the standard error of the mean.

### 7.3. Memory as a function of confidence

The results from the HLM analysis are presented for targets, related lures, and unrelated lures in Fig. 5A, B, and C, respectively. In addition to several main effects, this analysis resulted in a three-way item  $\times$  value  $\times$  confidence interaction ( $\beta = .07, p = .007$ ). To better understand the nature of this interaction, additional analyses were conducted separately for each item type. A value  $\times$  confidence interaction was found for targets ( $\beta = .22, p < .001$ ) and related lures ( $\beta = .18, p = .02$ ), but not unrelated lures ( $\beta = .07, p = .15$ ). For both targets and related lures, participants responded “yes” more often to words on the high-value test than the low-value test for high-confident responses, but responded “no” more often to words on the high-value test than the low-value test for low-confident responses. For unrelated lures, only a main effect of value was found ( $\beta = -.11, p = .05$ ), such that false alarms on the high-value test were reduced compared with those on the low-value test.

These results are consistent with the hypothesis that retrieval-monitoring processes were engaged (on average) across each value test, but also interacted with the level of confidence for items within each test. Specifically, across all item types (targets, related lures, and unrelated lures), retrieval monitoring was most effective for the high-value test than the low-value test when confidence was low, potentially when the fewest or least vivid details could be retrieved. In contrast, for high-confidence items (when retrieval monitoring was least likely to be engaged), responses appeared most liberal for high-value words than low-value words, thus exhibiting the traditional value effect on memory (average log odds of making a “yes” response averaged across targets and related lures were .47, .73, and 1.04 for low, medium, and high values respectively). This value-effect reversal was expected if assuming that retrieval monitoring was less likely to be engaged for high-confidence



**Fig. 4.** Memory and confidence results for Experiment 3. Mean proportion of source attributions to targets and source misattributions to related and unrelated lures (Panel A) and mean confidence ratings for targets, related lures, and unrelated lures (Panel B) as function of test (low, medium, and high value test). Error bars represent the standard error of the mean.

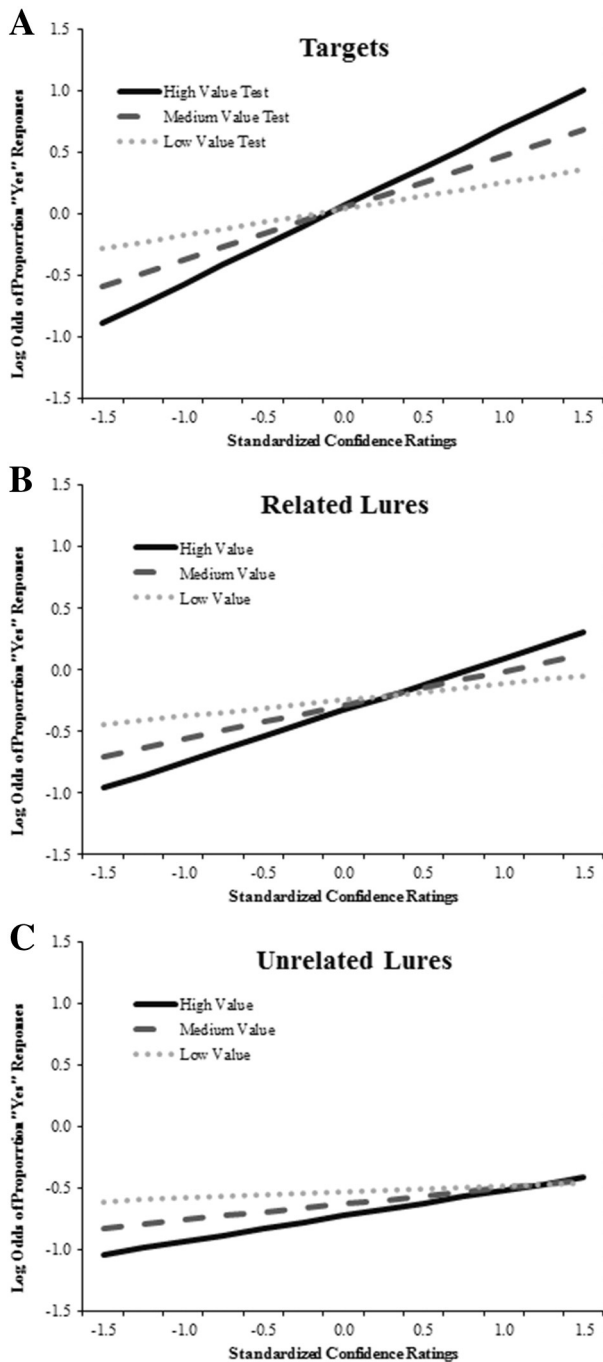
than low-confidence responses, thus revealing the effect of semantic activation or gist processes on memory, especially for related lures.

It should be noted that although the criterial recollection tests encourage the engagement of retrieval-monitoring processes (on average), clearly confidence level still moderated the degree to which people did actually engage in retrieval monitoring. Interestingly, confidence did not significantly moderate the degree of retrieval monitoring for unrelated lures, which are less likely to be associated with different levels of vividness/confidence (i.e., they are more homogenous). Rather, unrelated lures were more likely to reveal the effectiveness of retrieval monitoring on the high-value test (and was unopposed by associative activation/gist processing).

## 8. General discussion

The present study investigated the role that information importance plays during memory search and evaluation processes, known as retrieval monitoring. Effective retrieval monitoring is crucial to prevent false memories (e.g., Gallo, 2013; Johnson et al., 1993; Roediger et al., 2001), but how personal goals and motivations affect these processes had not been systematically explored. Understanding the impact of information importance on memory retrieval is critical to situate general principles of retrieval monitoring into a context more pertinent and reflective of our daily lives—one in which not all information is equally important to remember. In three experiments, participants were asked to selectively remember lists of words associated with higher-point values, and retrieval instructions were manipulated to either encourage or discourage the use of retrieval-monitoring processes while keeping encoding conditions constant. We found that encoding different levels





**Fig. 5.** Hierarchical linear modeling results for Experiment 3. Log odds of proportion “yes” responses as a function of confidence and value for targets (Panel A), related lures (Panel B), and unrelated lures (Panel C). Error bars represent the standard error of the mean.

of value impacted the quality of retrieved information, which then affected when retrieval-monitoring processes were employed and how effective those processes were in reducing false memories.

### 8.1. Monitoring is more effective for important information

Retrieval monitoring often is quite effective following the encoding of distinctive information (for review, see Schacter & Wiseman, 2006). It has been proposed that distinctive information is accompanied by the belief that details associated with that information should be particularly well remembered (i.e., heightened memorial expectations), which results in adjusting the type of details one uses to evaluate

memory (i.e., an increased reliance on item-specific details). This adjustment then reduces the susceptibility to memory illusions stemming from misrecollections or gist. Furthermore, it has been shown that these heightened memorial expectations, even in the absence of differences in distinctiveness, are sufficient to reduce false memories via monitoring processes (McDonough & Gallo, 2012). Previous findings have shown that more important information is expected to be better remembered (e.g., Festini et al., 2013; Friedman & Castel, 2011; Kassam et al., 2009; Soderstrom & McCabe, 2011), suggesting that more important information may lead to more effective retrieval monitoring. Consistent with this distinctiveness account, Experiments 1 and 3 showed that when retrieval monitoring was encouraged across all values, people more effectively monitored retrieval for information associated with higher relative to lower value as evidenced by the reductions in source misattributions. These reductions were particularly strong for unrelated items that were not susceptible to associative or gist processing that opposed the effectiveness of retrieval monitoring.

These findings have implications for situations in which motivation drives remembering, including instances of eyewitness memory. For example, in 2007, Vice-Presidential Chief of Staff I. Lewis “Scooter” Libby was indicted for leaking the covert identity of a Central Intelligence Agency (CIA) officer. On the witness stand, Libby claimed he could not remember what he previously told government officials or reporters regarding the identity of the CIA officer, despite the fact that this information was clearly important to know. Based on this event, Kassam et al. (2009) empirically tested whether knowing that information was important before or after learning the information would affect memory. They found that while people expected that knowing information both before and after learning should impact memory for the number of recallable facts, only knowing information before learning impacted recall performance. The present study builds on these findings in two ways. First, we showed that motivational processes can affect retrieval (as well as encoding). While our study did not manipulate when information was deemed important (i.e., before or after the study phase), simply believing that memory is better for more important information at retrieval might affect retrieval-monitoring processes, especially for non-criterial information (e.g., McDonough & Gallo, 2012). Second, we have shown that motivational processes also affect the creation (see also, Bui et al., 2013) and rejection of false memories. Taken together, motivation affects encoding processes, such as increasing the depth of processing engaged on each item, and retrieval processes such as when to engage monitoring and how effective monitoring is for each item.

### 8.2. Monitoring is less likely to be engaged for important information

While retrieval monitoring might be more *effective* for more important information, these processes must be first *engaged* to benefit from the additional search and evaluation processes. Under some contexts, people may not want or feel the need to engage in effortful retrieval monitoring. In Experiment 2, the degree of retrieval monitoring was manipulated using different types of instructions: inclusion instructions (i.e., endorse any item related to those previously seen as old) or standard instructions (i.e., endorse only previously seen items as old). Inclusion instructions should render monitoring processes less useful, and thus people should be less likely to engage retrieval monitoring as extensively. This manipulation affected responses to less important information (low-value words) more than more important information (high-value words), suggesting that retrieval monitoring for item-specific details was not occurring frequently or as extensively for important information. This finding has significant implications because information that we find most important is presumably what we also would like to be most accurate. But in stark contrast, it is more important information that is accompanied by the most false memories! While these false memories might be adaptive in some circumstances (cf. Bui et al., 2013), in other circumstances such as in eyewitness testimonials, accurate memory is needed most.

Why would people be less likely to monitor more important information compared with less important information? Less important information might be associated with fewer or less vivid details, thus decreasing confidence in the memory. This decreased confidence might then serve as a catalyst to engage additional memory search and evaluation processes. In contrast, information associated with higher value is likely more vividly recollected, thus increasing confidence in the memory (for both targets and lures). To the extent that people are very confident in the accuracy of their memory, they may not feel the need to engage in retrieval-monitoring processes to the same extent for higher relative to lower value information. Indeed, results from Experiment 3 indicated that higher-value items were retrieved on average with more confidence than lower-value items, further supporting this explanation. Ultimately, the current study suggests that the effects of value on retrieval monitoring are largely moderated by confidence level on a trial-by-trial basis. That is, across item type (targets or lures), across test type (standard recognition or criterial recollection), and across value type (low or high), low confidence is the key factor that determines when retrieval monitoring is engaged, but it is the qualitative characteristics of a class of items that determine the effectiveness of those retrieval-monitoring processes.

These ideas are consistent with generate-and-recognize models of memory (e.g., Jacoby & Hollingshead, 1990) in which participants first attempt to recall details and then engage in recognition and monitoring processes if the initially recalled details were not vividly or fluently retrieved. While the generate-and-recognize models primarily focus on memory recall rather than recognition, evidence from recognition studies supports this model. For instance, after reading a list of non-famous names, people were more likely to later judge the familiar names as famous in a fame-judgment task (Jacoby, Kelley, et al., 1989; Jacoby, Woloshyn, et al., 1989). The authors argued that the familiar names were retrieved fluently, and so retrieval-monitoring processes were not likely to be engaged. However, when encouraged to search memory for item-specific details, participants reduced their fame judgment errors. The flipside to this argument is that information that is retrieved less vividly or fluently is more likely to be monitored. Indeed, responses accompanied by lower confidence judgments (likely due to less vivid or fluent memories) also are associated with increased retrieval monitoring (e.g., Rotello & Macmillan, 2007). These ideas receive further support from Experiment 3, in which low-confidence responses (on a trial-by-trial basis) were associated with fewer endorsements of an event as having occurred (i.e., more “new” responses) compared with high-confidence responses, indicative of increased retrieval monitoring.

However, the false-fame studies investigated memory and the spontaneous use of retrieval monitoring (or lack thereof) in the context of a non-memory task (i.e., famous/nonfamous judgments). In the present experiments, memory was assessed in the context of an explicit memory task, and yet retrieval monitoring still varied across items, suggesting that one cannot assume that people will always engage in retrieval monitoring. Spontaneous monitoring may be even less likely to occur in the context of everyday life in which people have limited cognitive resources or time to deploy retrieval-monitoring processes (i.e., people are cognitive misers; Kahneman, 2003; Taylor, 1981). On the other hand, people might always be monitoring memory, but the types of details depend on the context of the situation. One possibility is that people may frequently monitor memory for the gist or familiarity of the memory rather than item-specific details of the memory, such as on a standard recognition memory test. Thus, the assumption that monitoring is reduced for more important information may be particular to monitoring item-specific details, which are most likely to aid in the reduction of false memories.

### 8.3. An alternative explanation: impoverished relational processing

We have interpreted the findings in the present study within a retrieval-monitoring framework, but one might ask whether we need

to invoke retrieval monitoring at all, or whether encoding differences as a function of value are sufficient to explain our findings. In the context of the current study, one might argue that we found a “more-is-less” effect (i.e., more value leads to less accurate memories) that were due to more relational processing for high-value words and impoverished-relational processing for low-value words. Studies have tested whether this type of “more-is-less” effect is due to different levels of impoverished-relational processing, retrieval monitoring, or a combination of both by using the same types of standard-memory instructions or inclusion (meaning-based) instructions as in Experiment 2. Sometimes this instruction manipulation has not entirely eliminated the “more-is-less” effect (e.g., Hege & Dodson, 2004; Hunt et al., 2011), suggesting that both accounts may play a role in affecting false memories. Other times, however, the “more-is-less” effect was completely eliminated (e.g., Pierce et al., 2005), suggesting a strong role for retrieval-monitoring processes. In the present study, we showed that this instruction manipulation completely eliminated the effects of value on false memories (and true memories), suggesting that the impoverished-relational account cannot explain the value effects on memory, but rather retrieval-monitoring processes play a critical role.

Even when using an inclusion-instruction manipulation, retrieval monitoring effects can sometimes be hard to assess in the context of the DRM paradigm. While studies using the DRM paradigm have led to many insights on the creation of false memories, associative activation or gist processing potentially masks the use of retrieval-monitoring processes that are recruited to help reduce potential false memories (cf., Gallo, 2010, 2013). One strength of the present study is that we used different paradigms (DRM and non-DRM) to reveal not only that retrieval-monitoring processes are masked by semantic activation or gist processing (as in Experiment 3 for targets and related lures), but also that retrieval monitoring of item-specific details may not even be engaged to the same extent under standard memory test conditions (as in Experiment 2). Furthermore, Experiments 1 and 3 found that false memories for non-studied lures (which did not undergo relational processing) varied as a function of value, also suggesting that information value has an independent effect on retrieval-monitoring processes.

## 9. Conclusion

The present study illustrated the complex interplay between those processes leading to increases in false memories (semantic activation and/or gist creation) and those processes leading to decreases in false memories (retrieval monitoring). One key factor influencing this dynamic interplay is the importance of information. Because some information may be perceived as more important than others, people are motivated to elaborately encode some pieces of information at the expense of others. Ironically, these motivational factors can lead to less retrieval monitoring, and in turn, increase the susceptibility to false memories. While we cannot afford to expend our cognitive resources to closely inspect every memory, the present study suggests that the added effort to do so can offset potential false memories for information we deem important.

## Acknowledgments

We thank David Gallo, Sara Festini, and the reviewers for helpful comments at various stages of this work.

## References

- Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, 138, 432–447.
- Asp, E.W., & Tranel, D. (2012). False tagging theory: toward a unitary account of prefrontal cortex function. In D.T. Stuss, & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 383–416). New York, NY: Oxford University Press.
- Brainerd, C.J., & Reyna, V.F. (1998). When things that never happened are easier to “remember” than things that did. *Psychological Science*, 9, 484–489.

- Bui, D.C., Friedman, M.C., McDonough, I.M., & Castel, A.D. (2013). False memory and importance: Can we prioritize encoding without consequence? *Memory & Cognition*, *41*, 1012–1020.
- Busey, T.A., Tunnicliff, J., Loftus, G.R., & Loftus, E.F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*, 26–48.
- Castel, A.D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. *The Psychology of Learning and Motivation*, *48*, 225–270.
- Castel, A.D., Benjamin, A.S., Craik, F.I.M., & Watkins, M.J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078–1085.
- Castel, A.D., Farb, N., & Craik, F.I.M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition*, *35*, 689–700.
- Cohen, M.S., Rissman, J., Harbert, E., Castel, A.D., & Knowlton, B.J. (2013). Enhancement of recognition memory by point value is associated with increased recollection. *Paper presented at the 53rd Annual Meeting of the Psychonomic Society, Toronto, Canada* (November).
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.
- Dodson, C.S., & Johnson, M.K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology*, *706*, 541–557.
- Festini, S.B., Hartley, A.A., Tauber, S.K., & Rhodes, M.G. (2013). Assigned value improves memory of proper names. *Memory*, *21*, 657–667.
- Friedman, M.C., & Castel, A.D. (2011). Are we aware of our ability to forget? Metacognitive predictions of directed forgetting. *Memory & Cognition*, *39*, 1448–1456.
- Gallo, D.A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848.
- Gallo, D.A. (2013). Retrieval expectations affect false recollection: Insights from a criterial-recollection task. *Current Directions in Psychological Science*, *22*, 316–323.
- Gallo, D.A., Kensinger, E.A., & Schacter, D.L. (2006). Prefrontal activity and diagnostic monitoring of memory retrieval: fMRI of the criterial-recollection task. *Journal of Cognitive Neuroscience*, *18*, 135–148.
- Gallo, D.A., McDonough, I.M., & Scimeca, J.M. (2010). Dissociating source memory decisions in the prefrontal cortex: fMRI of diagnostic and disqualifying monitoring. *Journal of Cognitive Neuroscience*, *22*, 955–969.
- Gallo, D.A., Meadow, G.N., Johnson, E.L., & Foster, K.T. (2008). Levels of processing and distinctiveness: Evidence from the criterial-recollection task. *Journal of Memory and Language*, *58*, 1095–1111.
- Gallo, D.A., Weiss, J.A., & Schacter, D.L. (2004). Reducing false recognition with criterial-recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language*, *51*, 473–493.
- Hege, A.C., & Dodson, C.S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 787–795.
- Hennessee, J., Cohen, M.S., Castel, A.D., & Knowlton, B.J. (2014). Item value primarily enhances recollection in recognition memory. *Paper presented at the 54th Annual Meeting of the Psychonomic Society, Long Beach, California* (November).
- Hunt, R.R., & Einstein, G.O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *19*, 497–514.
- Hunt, R.R., & McDaniel, M.A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421–445.
- Hunt, R.R., Smith, R.E., & Dunlap, K.R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language*, *65*, 378–389.
- Jacoby, L.L., & Hollingshead, A. (1990). Toward a generate/recognize model of performance on direct and indirect tests of memory. *Journal of Memory and Language*, *29*, 433–454.
- Jacoby, L.L., Kelley, C.M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*, 326–338.
- Jacoby, L.L., Woloshyn, V., & Kelley, C.M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, *118*, 115–125.
- Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720. <http://dx.doi.org/10.1037/0003-066X.58.9.697>.
- Kassam, K.S., Gilbert, D.T., Swencionis, J.K., & Wilson, T.D. (2009). Misconceptions of memory: The Scooter Libby effect. *Psychological Science*, *20*, 551–552.
- Lindsay, D.S., & Johnson, M.K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349–358.
- Loftus, G.R., & Wickens, T.D. (1970). Effect of incentive on storage and retrieval processes. *Journal of Experimental Psychology*, *85*, 141–147.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*, 1–23.
- McDonough, I.M., & Gallo, D.A. (2008). Autobiographical elaboration reduces memory distortion: Cognitive operations and the distinctiveness heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1430–1445.
- McDonough, I.M., & Gallo, D.A. (2012). Illusory expectations can affect retrieval-monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 391–404.
- Multhaup, K.S., De Leonardis, D.M., & Johnson, M.K. (1999). Source memory and eyewitness suggestibility in older adults. *The Journal of General Psychology*, *126*, 74–84.
- Pierce, B.H., Gallo, D.A., Weiss, J.A., & Schacter, D.L. (2005). The modality effect in false recognition: Evidence for test-based monitoring. *Memory & Cognition*, *33*, 1407–1413.
- Reinitz, M.T., Peria, W.J., Séguin, J.A., & Loftus, G.R. (2011). Different confidence-accuracy relationships for feature-based and familiarity based memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 507–515.
- Reyna, V.F., & Brainerd, C.J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.
- Roediger, H.L., Balota, D.A., & Watson, J.M. (2001). Spreading activation and the arousal of false memories. In H.L. Roediger, J.S. Nairne, I. Neath, & A.M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association Press.
- 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.
- Rotello, C.M., & Macmillan, N.A. (2007). Response bias in recognition memory. *Psychology of Learning and Motivation*, *48*, 61–94.
- Schacter, D.L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1–24.
- Schacter, D.L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language*, *35*, 319–334.
- Schacter, D.L., & Wiseman, A.L. (2006). Reducing memory errors: The distinctiveness heuristic. *Distinctiveness and Memory* (pp. 89–107). New York, NY, US: Oxford University Press.
- Scimeca, J.M., McDonough, I.M., & Gallo, D.A. (2011). Quality trumps quantity at reducing memory errors: Implications for retrieval monitoring and mirror effects. *Journal of Memory and Language*, *65*, 363–377.
- Smith, R.E., Hunt, R.R., & Gallagher, M.P. (2008). The effect of study modality on false recognition. *Memory & Cognition*, *36*, 1439–1449.
- Soderstrom, N.C., & McCabe, D.P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1236–1242.
- Spaniol, J., Schain, C., & Bowen, H.J. (2014). Reward-enhanced memory in younger and older adults. *The Journal of Gerontology: Psychological Sciences*, *69*, 730–740.
- Taylor, S.E. (1981). The interface of cognitive and social psychology. In J.H. Harvey (Ed.), *Cognition, social behavior, and the environment* (pp. 189–211). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.