Individual Differences in Disqualifying Monitoring Underlie False Recognition of Associative and Conjunction Lures

B. Hunter Ball[1] · Matthew K. Robison[1] · Allison Coulson[2] · Gene A. Brewer[2]

Abstract

The current study leveraged experimental and individual differences methodology to examine whether false memories across different list-learning tasks arise from a common cause. Participants completed multiple false memory (associative and conjunction lure), working memory (operation and reading span), and source monitoring (verbal and picture) tasks. Memory discriminability in the associative and conjunction tasks loaded onto a single (general) factor and were unaffected by warnings provided at encoding. Consistent with previous research, source-monitoring ability fully mediated the relation between working memory and false memories. Moreover, individuals with higher source monitoring-ability were better able to recall contextual information from encoding to correctly reject lures. These results suggest that there are stable individual differences in false remembering across tasks. The commonality across tasks may be due, at least in part, to the ability to effectively use disqualifying monitoring processes.

Keywords False memory · Working memory · Source monitoring · individual differences · Rejection strategies

Examination of false memories provides insights into the processes that underlie the organization of memories, with important implications in legal (Loftus, 1975, 2019), clinical (Loftus & Pickrell, 1995; Pope, 1996), and interpersonal contexts (Brashier & Schacter, 2020; Meade & Roediger, 2002). For example, memory errors associated with eyewitness misidentifications are the leading cause of wrongful convictions that have been later overturned by DNA evidence (West & Meterko, 2015). While the ubiquity of false memory research has undoubtedly contributed to understanding the principles underlying memory retrieval, the variety of false memory paradigms has made it difficult to determine whether similar memory monitoring mechanisms operate across different contexts. The purpose of the current paper is to leverage experimental and individual differences methodology to determine

which cognitive processes underlie false recognition of associative and conjunction lures and the extent to which these processes operate across seemingly different list-learning paradigms.

Two widely used methods to examine false memories are the Deese–Roediger–McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995) and the memory conjunction (MC) paradigm (Lampinen et al., 2004; Odegard & Lampinen, 2005). The typical DRM finding in recognition memory is that after studying a list of associatively or semantically related words (e.g., *bed, rest, tired*), people erroneously claim that a nonpresented critical lure (e.g., *sleep*) was originally studied. The general finding in the MC paradigm is that after studying a set of compound words (e.g., *nosebleed, skydive*), people often erroneously claim that a nonpresented critical lure composed of features of previously presented items (e.g., *nosedive*) was originally studied. The primary difference between these paradigms is that DRM errors primarily arise through semantic or associative similarities between studied items and lures, whereas MC errors arise via phonological or perceptual similarities (Matzen & Benjamin, 2009; Matzen et al., 2011; Odegard et al., 2005). In both instances, however, it is thought that the high degree of overlap between studied items (e.g., *bed*; *nosebleed*) and critical lures (e.g., *sleep*; *nosedive*) causes participants to mistake lure items as having previously been experienced. The question here is

✉ B. Hunter Ball
Hunter.Ball@uta.edu

1 Department of Psychology, University of Texas at Arlington, Arlington, TX 76019, USA

2 Department of Psychology, Arizona State University, Arlington, TX, USA

🖉 Springer

whether a general monitoring mechanism underlies false remembering across these different tasks.

Dual-process theories of memory posit that recognition memory occurs via either automatic (familiarity) or controlled (recollection) processes (Jones & Jacoby, 2001). False memory paradigms are successful in eliciting memory errors because the similarity between lures and previously studied items often invokes a feeling of lure familiarity in the absence of recollection. This idea is central to two of the more prominent theories of false remembering.

Activation-monitoring theory suggests studying a list of related items implicitly activates the critical lure via automatic spreading activation during encoding (Anderson, 1983; Collins, & Loftus, 1975; Gallo, 2006; Gallo & Roediger, 2002; Roediger & McDermott, 1995), whereas the fuzzy trace theory argues that participants extract the overall theme, or gist, of the study lists (Brainerd & Reyna, 2001, 2004; Brainerd et al., 2001). At test the activated lure, or gist trace, produces a strong feeling of familiarity. To reject these lures, participants must decide whether the item was previously seen (e.g., *bed*; *nosebleed*), it was merely imagined (e.g., *sleep*), or it contains some but not all elements from study (e.g., *nosedive*; Johnson et al., 1993). Both theories argue that errors occur when recollective retrieval processes fail at test. Although the exact retrieval mechanisms slightly differ across theories (source monitoring vs. recollection rejection), for the purposes of the current study we will use the term "monitoring" to describe the lure rejection process.

Gallo describes two different ways in which monitoring can be implemented to reduce false memories (Gallo, 2004, 2006, 2010). *Disqualifying monitoring* is a rule-based strategy in which retrieval of contextual information from study disqualifies a familiar lure as having been previously encountered (Hintzman & Curran, 1994; Jacoby, 1991), whereas *diagnostic monitoring* is a heuristic-based strategy where the quality of the retrieved details is compared to an otherwise expected amount (Johnson et al., 1993; Mitchell & Johnson, 2000, 2009). For example, using a disqualifying monitoring strategy a lure may be rejected by recalling specific items from study ("I remember seeing skydive, so I could not have seen nosedive"). In contrast, when using a diagnostic monitoring strategy, a lure may be rejected based on an inference that the item would have been remembered if actually studied ("I would have remembered seeing nosedive because I recently flew in a plane that had a rough landing"). While both are considered a type of recollective monitoring, they differ in the *type* of retrieved information used to reject items.

Research in the DRM paradigm supports the idea that both types of monitoring can reduce errors. For example, warning participants of the illusion prior to encoding and presenting lures in a to-be-excluded list increases diagnostic monitoring (Gallo et al., 2006; Neuschatz et al., 2003). Presumably these manipulations help in identifying the theme during encoding (e.g., *sleep*), which subsequently allow participants to disqualify it as having be presented. Additionally, being able to recall all items from the studied list (e.g., *bed, rest, tired*) can disqualify the lure (Gallo, 2004). Alternatively, manipulations that increase the distinctiveness of studied items (e.g., pictures versus words) increases diagnostic monitoring by allowing participants to reject lures that do not have distinctive representations (Arndt & Reder, 2003; Benjamin, 2001; Dodson & Schacter, 2001; Hege & Dodson, 2004; Smith & Hunt, 1998). Research in the MC paradigm has largely focused on disqualifying monitoring. Several studies have assessed phenomenological reports at retrieval, finding that following rejection of compound lures (e.g., *nosedive*) participants often report being able to recall related words from study (e.g., *nosebleed*; *skydive*; Arndt & Jones, 2008; Leding & Lampinen, 2009; Odegard et al., 2005). Warnings at retrieval also cause participants to monitor memory more carefully for the studied associates (Lampinen et al., 2004). Note that this is slightly different from the DRM paradigm in which warnings primarily only reduce errors when given at encoding (Gallo et al., 1997; Gallo et al., 2001; McDermott & Roediger, 1998). It has also been shown that distinctive processing at encoding reduces false alarms (Arndt & Jones, 2008; Lloyd, 2007). However, differing from the DRM paradigm, this appears to be driven by decreased familiarity for lures rather than increased diagnostic monitoring (Lloyd, 2013). Together these findings highlight monitoring as an important construct, although what information is used to reject lures may differ across paradigm types.

Generality of monitoring

A central question for theory development is whether a common memory monitoring mechanism operates across different false-memory tasks (Gallo, 2010). One way to address this issue is to assess individual differences in performance across multiple false memory paradigms. Assuming there is a common mechanism (e.g., monitoring), individuals with the propensity to false alarm on one type of task (e.g., DRM) should have higher false alarm rates on another task (e.g., MC). Generally consistent with this idea, prior research has shown that DRM errors are positively correlated with false memories for words and pictures (Lövdén, 2003; Unsworth & Brewer, 2010), autobiographical memories (e.g., Meyersburg et al., 2009; Platt et al., 1998), and misinformation (Zhu et al., 2013). However, other studies have not always found these associations (Falzarano & Siedlecki, 2019; Monds et al., 2017; Nichols & Loftus, 2019; Ost et al., 2013; Patihis et al., 2018; Salthouse & Siedlecki, 2007; Wilkinson & Hyman, 1998). The inconsistencies across studies could indicate that (a) there is no general monitoring mechanism, (b) there is a common mechanism that operates differently across

paradigms (e.g., diagnostic versus disqualifying monitoring), or (c) there is a common monitoring mechanism, but it is not a stable individual difference (Patihis et al., 2018; but see Unsworth & Brewer, 2010). Notably, many of these studies used only a single indicator of false memory across different paradigms. As described later, this is not ideal in the context of individual differences because any observed effects (or lack thereof) may simply be an artifact of shared method variance or measurement error.

Another approach to examining the generality of monitoring processes is to examine whether certain cognitive abilities (e.g., monitoring) are related to individual differences in false remembering. One cognitive ability measure commonly assessed in false memory studies is working memory. Working memory refers to the attention processes involved in active maintenance of task-relevant information in the face of distraction (Kane & Engle, 2003) and the memory processes associated with the retrieval of information from long-term memory that has been momentarily displaced from focal awareness (Unsworth & Engle, 2007). Research has shown that individuals with higher working memory capacity show fewer errors in the DRM paradigm (Holden et al., 2020; Lövdén, 2003; McCabe & Smith, 2002; Peters et al., 2007; Unsworth & Brewer, 2010; Watson et al., 2005), MC paradigm (Leding, 2012), misinformation paradigm (Jaschinski & Wentura, 2002; Parker et al., 2008; Zhu et al., 2010), and for missing events (Gerrie & Garry, 2007). The general interpretation of these findings is that that high working memory participants are better able to monitor retrieval to determine the source of a memory (e.g., "did I study this or merely imagine it"; but see Watson et al., 2005). While this interpretation is theoretically appealing, it is admittedly a relatively indirect way of assessing the mechanisms underlying false remembering.

Unsworth and Brewer (2010) more directly assessed the role of monitoring in DRM errors by assessing multiple measures of source memory in addition to working memory tasks.

Additionally, they had participants perform the DRM task along with several standard delayed free recall and paired associate learning tasks. They found that errors in the DRM, free recall, and paired associates tasks loaded onto a single false memory factor. This means that if a participant had high errors on one task, they were more likely to have high errors on another. Critically, they also found that the relation between working memory and false memory was fully mediated by source-monitoring ability. That is, the reason high working memory individuals had fewer memory errors is because they were better able to monitor for the origin of a memory. This suggests that there may be a common memory monitoring mechanism across multiple word list tasks. It is not entirely clear, however, if the false memory reduction was due to better disqualifying or diagnostic monitoring for high ability participants.

A study by Leding (2012) partially addressed the issue of what type of monitoring high ability participants use. Participants performed an MC task with and without warnings are were to indicate what strategy they used to reject lures. High working memory participants showed fewer MC errors regardless of whether or not they were warned of the illusion at encoding. These results are consistent with the findings of Unsworth and Brewer (2010) indicating that high ability participants better monitor retrieval outputs. Critically, these individuals were also more likely to engage in disqualifying monitoring. That is, they were better able to retrieve the compound words from study to reject the lure at test. However, because diagnostic monitoring was not assessed in this study, it is not entirely clear whether high and low ability participants differ in the efficacy of disqualifying monitoring or if they rely on qualitatively different types of information (i.e., disqualifying versus diagnostic details) to reject lures.

Current study

The purpose of the current study was to determine whether a general memory monitoring process might underlie false recognition of associative and conjunction lures that have seemingly disparate processing demands. Participants performed multiple false memory word list tasks (DRM and MC) with and without warnings at encoding, along with several working memory and source monitoring tasks. During the false-memory tasks, participants also reported the strategies used to reject lures (e.g., diagnostic monitoring, disqualifying monitoring, or lacks familiarity). Using this methodology, we hoped to improve on several theoretical and methodological shortcomings of prior experiments. First, using multiple performance indicators allows for the use of latent variable modeling. This approach is useful because it controls for measurement error while testing different theoretical predictions of the relation between false memory, working memory, and source monitoring. Second, the inclusion of the source memory tasks allows for a more direct assessment of the role of monitoring in false memory. Third, assessing rejection strategies allows for the determination of whether high and low ability participants use quantitatively or qualitatively different information to reduce false memories.

Regarding the generality of a memory monitoring mechanism, previous research is mixed. While in some cases false remembering on one task (e.g., DRM) is correlated with false memory on another task (e.g., misinformation; Lövdén, 2003; Meyersburg et al., 2009; Platt et al., 1998; Unsworth & Brewer, 2010; Zhu et al., 2013), other times it is not (Monds et al., 2017; Ost et al., 2013; Patihis et al., 2018; Salthouse & Siedlecki, 2007; Wilkinson & Hyman, 1998). We test the

generality claim by comparing the fit of a theoretical model in which all false-memory tasks load onto a common factor (general model) to a model in which the different task types load onto separate factors (task-specific model). We anticipated that the general model would provide a more adequate account of the data, suggesting that there may be a general memory monitoring mechanism that contributes to both tasks. Regarding individual differences in performance, prior research suggests that high working memory participants may be better able to monitor retrieval to determine the source of a memory (e.g., Gerrie & Garry, 2007; Leding, 2012; McCabe & Smith, 2002; Unsworth & Brewer, 2010). We explicitly tested this claim by assessing whether the relation between working memory and false memory would remain after controlling for source memory. We anticipated that source memory would fully mediate this relation (Unsworth & Brewer, 2010). Finally, regarding what type of monitoring is used by individuals of differing ability, there simply is not enough research on the topic to make any clear predictions (Leding, 2012). To address this concern, we assessed the relation between source monitoring and the different rejection strategies. We anticipated that high ability participants would more often use disqualifying monitoring, whereas low ability participants would rely more on diagnostic monitoring.

Method

All research reported herein was conducted using appropriate ethical guidelines and was approved by the Institutional Review Board at Arizona State University. We report how we determined our sample size, all data exclusions, and all manipulations.

Participants and design

For the no-warning and warning conditions, respectively, 247 and 205 undergraduate participants were recruited from the Arizona State University participant pool. These sample sizes were selected based on previous research using a similar methodology (Unsworth & Brewer, 2010) and recommendations that at least 150 participants are needed to obtain stable and reliable correlations (Schönbrodt & Perugini, 2013). Data were excluded from six participants that were identified as multivariate outliers using Mahalonbis distance estimates. The final dataset consisted of 245 and 201 participants, respectively, in the no warning and warning condition. All participants were native English speakers and received course credit for their participation. Participants were tested in group laboratory sessions (from one to seven participants) lasting approximately 2 hours.

Cognitive battery

Participants completed two versions of DRM, MC, working memory, and source memory tasks.[1] The DRM and MC tasks were nearly identical across conditions, except that in the warning condition participants were warned prior to study about the presence of critical lures at test. The working memory and source monitoring tasks were identical across conditions. Learning was intentional in all tasks.

Working memory tasks

Operation span (SPANo) Participants solved a series of math operations while trying to remember a set of unrelated letters (F, H, J, K, L, N, P, Q, R, S, T, Y). Participants were required to solve a math operation, and after solving the operation they were presented with a letter for 1 s. Immediately after the letter was presented the next operation was presented. Three trials of each list length (3–7) were presented, with the order of list-length varying randomly. At recall, letters from the current set were recalled in the correct order by clicking on the appropriate letters (see Unsworth et al., 2005, for more details). Participants received three sets (of list-length two) of practice. For all of the span measures, items were scored if the item was correct and in the correct position. The score is the proportion of correct items in the correct position.

Reading span (SPANr) Participants were required to read sentences while trying to remember a set of unrelated letters (F, H, J, K, L, N, P, Q, R, S, T, Y). For this task, participants read a sentence and determined whether the sentence made sense or not (e.g., "The prosecutor's dish was lost because it was not based on fact"). Half of the sentences made sense while the other half did not. Nonsense sentences were made by simply changing one word (e.g., "dish" from "case") from an otherwise normal sentence. Participants were required to read the sentence and to indicate whether it made sense or not. After participants gave their response, they were presented with a letter for 1 s. At recall, letters from the current set were recalled in the correct order by clicking on the appropriate letters. There were three trials of each list-length with list-length ranging from 3–7. The same scoring procedure as Ospan was used.

Source Monitoring Tasks

Gender source recognition (SMg) Participants heard words (30 total words) in either a male or a female voice. Participants were explicitly instructed to pay attention to both the word as well as the voice the word was spoken in. At test participants

---

[1] As part of a larger battery for an unrelated purpose, participants also completed several measures of cued recall, free recall, and recognition memory.

were presented with 30 old and 30 new words and were required to indicate whether the word was new or old and, if old, what voice it was spoken in via key press. Participants had 5 s to press the appropriate key to enter their response. A participant's score was the proportion of correct responses.

Picture source recognition (SMp) Participants were presented with a picture (30 total pictures) in one of four different quadrants on screen for 1 s. Participants were explicitly instructed to pay attention to both the picture as well as the quadrant in which it was located. At test, participants were presented with 30 old and 30 new pictures in the center of the screen. Participants indicated whether the picture was new or old and, if old, what quadrant it was presented in via key press. Participants had 5 s to press the appropriate key to enter their response. A participant's score was the proportion of correct responses.

False-memory task materials

DRM word (DRMw) Participants studied 8 lists of 8 words for a total of 64 words. Each list was composed of semantically associated words that were all related to a critical word. These lists were taken from Alakbarova et al. (2021; Experiment 2), which were adapted from Thomas & Sommers (2005). The eight critical words included *city, smoke, soft, sweet, foot, mountain, thief,* and *river*. Study presentation was blocked by list, the order of which was randomized for each participant. The order in which the eight words within each list were presented was random. At test, participants were shown a total of 80 selected words (40 old and 40 new items) in a random order (see also Thomas & Sommers, 2005). For the primary analyses, the old items consisted of five words (e.g., *rest, dream, bed*) from each of the eight lures sets, and the new items consisted of eight critical lures (e.g., *sleep*) and eight unrelated new items (e.g., *horse*). The five words from each list were randomly selected preexperimentally and were the same for each participant. The d' lure value was calculated as a function of hits and lure false alarms, whereas d' new value was calculated as a function of hits and unrelated new false alarms. To allow for the calculation of d' values when hit or false alarm rates equaled zero or one, the Hautus (1995) correction was used for signal detection analyses. There were also 24 other new words that were not included in the analyses, which included four new theme words (e.g., *fork*) and five words associated with each of these themes (e.g., *spoon, knife*). The five words from the new lists were also randomly selected preexperimentally and were the same for each participant.

DRM sentence (DRMs) Participants studied eight lists of 8 sentences for a total of 64 sentences. In each list, the associate was embedded as the last word of the sentence and all

revolved around a lure theme (e.g., *sleep* lure: "After work he laid down in bed"; "She had a frightening dream"). The eight critical lures related to the sentences were *cup, trash, window, doctor, cold, chair, sleep,* and *needle* from the convergent condition of Alakbarova et al. (2021; Experiment 2). The test procedure was identical to that of the DRMw task.

MC word (MCw) In this task, participants were presented with 48 semantically unrelated words (e.g., *ragtime, tumbleweed; snowman, football*) one at a time in a random order. Two of the compound words were related to one of 24 sets of semantically unrelated MC lures (e.g., *ragweed; snowball*) at study. These stimuli were provided by Matzen and Benjamin (2009). At test, participants were presented with 48 total words (24 old and 24 new items) in a random order. For the primary analyses, the old items consisted of 24 compound words (e.g., *ragtime*; 12 parent one and 12 parent two items), and the new items consisted of 12 critical lures (e.g., *snowball*) and 12 unrelated new words (e.g., *bookbag*). The 12 critical lures came from the parent items not presented at test, meaning that there was no overlap between old and new items (i.e., both the compound word *ragtime* and the critical lure *ragweed* were not presented). The d' lure value was calculated as a function of hits and lure false alarms, whereas d' new value was calculated as a function of hits and unrelated new false alarms using the Hautus (1995) correction.

MC sentence (MCs) In this task, participants were presented with 48 semantically unrelated sentences containing compound words (e.g., *bandwagon, kickstand; catwalk, shellfish*) one at a time in a random order. Two of the sentences were associated with one of 24 sets of semantically unrelated MC lures (e.g., *bandstand; catfish*) at study. The associate was embedded as the last word of the sentence (e.g., *bandstand* lure: "After the team won the championship hundreds of new fands jumped on the bandwagon"; "The girl propped her new bicycle on its kickstand"). The test procedure was identical to that of the MCw task.

False-memory task procedure

For all four of the false-memory tasks, during study participants were presented with one stimulus at a time and instructed to subjectively rate the meaning of the words on a scale of 1–7 (1 = *very meaningless*, 7 = *very meaningful*). This orienting task was based on previous methodology from our laboratory where these materials were used (Alakbarova et al., 2021). They were told that their memory would later be tested for these words. At test, participants were presented with one stimulus at a time and asked to choose if it was "old" or "new." If determined "old," they had to further categorize it as (a) know it's familiar (i.e., familiarity), (b) remember details (i.e., recollection), or (c) guess. If determined "new,"

participants were instructed to classify it as either (a) lacks familiarity, (b) recalled study information to reject it (i.e., disqualifying monitoring), or (c) would have remembered (i.e., diagnostic monitoring). Participants were provided with clear examples on what each classification reflected and were instructed to choose the one that best fit their memory for the event. Test instructions introduced an approximately 2-minute delay between study and test.

The examples provided to participants for the different classifications of the rejection strategies were as follows. *Lacks familiarity*: "The word did not seem or feel familiar because it lacked familiarity, thus one believes that it was not studied. This is similar to telling Person X that you do not remember meeting him because he just 'doesn't look familiar to you' and that you do not recognize him," *Recall to reject*: "One may feel that they searched memory and recalled something about the study phase that lead them to believe the word was not studied. In this way, the test item can be rejected because one searches and may find similar items, but not the specific test item. For example, if the test word is 'train,' you may search your memory of the study phase and remember that you studied "plane." This leads you to believe that you did not study the similar word, 'train.'" *Would have remembered*: "One thinks that the word is so distinctive that if it was studied, it would have been remembered, and because it is not remembered it must not have been studied. For example, if your first name appeared on the test you might judge it as new because you feel that it would have been so distinctive to you that if you did study it, you would have remembered it."

The only difference between no warning and warning conditions was that in the former participants were told at encoding that their memory would be tested for the stimuli but were given no more information about the nature of the study or test materials. In the latter participants were additionally informed prior to study about the existence of critical lures. Following McCabe and Smith (2002), participants were given an example and then were instructed during study to try to identify the lures. They were also told that these lures would later appear on the test and they should avoid recognizing them. In the word and sentence DRM task participants were explicitly told that each list of eight trials (e.g., *bed, rest, tired*) was associated with one common word (e.g., *sleep*), and were instructed to try to determine what the common word was. In the word and sentence MC lure task participants were explicitly instructed that each of the study words (e.g., *tumbleweed, ragtime*) were compound words whose parts could be combined to form a conjunction word (e.g., *ragweed*), and were told to try to identify what the conjunction word was. Participants were warned prior to study that the DRM lures (e.g., *sleep*) and MC lures (e.g., *ragweed*) would be presented at test and that they should avoid calling them "old."

Data analysis

Matzen and Benjamin (2009) note that using standard false-alarm rates to critical lures as a measure of false memory is only appropriate when participants adopt similar criterion across conditions. If criterion differs but memory is the comparable across conditions, then the more appropriate false memory measure is sensitivity between old and lure items. In the current study, because participants studied words and sentences in both DRM and MC tasks, we anticipated differences in both memory and criterion. In such instances, they argue that the most appropriate false memory measure is the difference ($\Delta d'$) between d' new (from hits and unrelated new false alarms) and d' lure (from hits and critical lures false alarms) measures ($\Delta d' = d'$ new $- d'$ lure). The primary measure for all analyses in the current study is therefore the $\Delta d'$ value. A positive value indicates that the participant was better able to discriminate new unrelated items from critical lure items.

For the primary analyses, we use confirmatory factor analysis (CFA) to specify a theoretically derived model and compare this to the true variance-covariance matrix for the observed data (Kline, 2015). A chi-squared test is used to determine how well the specified model reproduces the observed data, with a nonsignificant value indicating a good fit. We additionally report several other goodness-of-fit indices: root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), nonnormed fit index (NNFI), and comparative fit index (CFI). The RMSEA and SRMR reflect the average squared deviation between the observed and reproduced covariances, whereas the NNFI and CFI compare the fit of the specified model to a baseline null model. After identifying the best fitting theoretical model through a series of chi-square differences tests (notated as $\Delta \chi^2$ in the results section), we use latent variable structural equation model (SEM) to assess unique contributions of cognitive ability (i.e., working memory and source monitoring) in predicting false memories. Missing data were accounted for using maximum likelihood estimation.

Results

Descriptive statistics for overall hit and false-alarm rates in the false-memory tasks and overall memory performance in the working memory and source monitoring task can be found in Table 1. All measures had acceptable values of kurtosis (skew < |3| and kurtosis < |8|; Kline, 2015). The primary dependent variable for all false-memory tasks was $\Delta d'$ (see Data

---

[0] We also assessed discriminability of lure items from unrelated items, which is considered a measure of false memory while controlling for response bias. These analyses produced a similar pattern of results to the $\Delta d'$ analyses.

Table 1 Descriptive statistics and reliability estimates for all measures

Condition	Task	DV	Mean	SD	Min	Max	Skew	Kurtosis	α
No Warning	DRMw	Hit	0.76	0.12	0.23	0.98	-1.29	1.87	0.81
		FA New	0.19	0.20	0.00	1.00	1.21	1.18	0.58
		FA Lure	0.56	0.21	0.13	1.00	-0.16	-0.58	0.45
	SM	Hit	0.70	0.12	0.40	0.95	-0.37	-0.01	0.67
		FA New	0.31	0.19	0.00	0.88	0.55	-0.07	0.53
		FA Lure	0.76	0.19	0.25	1.00	-0.67	-0.17	0.42
	DRMs	Hit	0.89	0.11	0.42	1.00	-1.54	2.70	0.71
		FA New	0.13	0.16	0.00	0.92	2.06	5.15	0.73
		FA Lure	0.28	0.21	0.00	1.00	0.77	0.05	0.70
	MCw	Hit	0.74	0.15	0.25	1.00	-0.71	0.03	0.70
		FA New	0.21	0.22	0.00	1.00	1.18	0.98	0.78
		FA Lure	0.41	0.20	0.08	1.00	0.36	-0.28	0.58
	MCs	SPANo	0.77	0.16	0.13	1.00	-1.07	1.22	0.98
		SPANr	0.66	0.21	0.01	1.00	-0.79	-0.03	0.98
	WM	SMg	0.53	0.16	0.12	0.92	0.01	-0.54	0.85
		SMp	0.65	0.22	0.03	0.98	-0.84	-0.09	0.95
Warning	DRMw	Hit	0.77	0.11	0.38	1.00	-1.19	1.80	0.76
		FA New	0.20	0.23	0.00	1.00	1.42	1.77	0.72
		FA Lure	0.55	0.23	0.00	1.00	-0.20	-0.64	0.57
	DRMs	Hit	0.70	0.13	0.25	0.95	-0.67	0.95	0.73
		FA New	0.31	0.19	0.00	1.00	0.89	0.87	0.53
		FA Lure	0.72	0.22	0.13	1.00	-0.56	-0.51	0.56
	MCw	Hit	0.87	0.12	0.17	1.00	-2.14	7.06	0.73
		FA New	0.13	0.17	0.00	1.00	2.06	4.88	0.78
		FA Lure	0.24	0.20	0.00	1.00	1.08	1.09	0.69
	MCs	Hit	0.76	0.15	0.00	1.00	-1.18	2.46	0.72
		FA New	0.20	0.21	0.00	0.83	1.17	0.67	0.77
		FA Lure	0.41	0.18	0.00	0.92	0.17	-0.59	0.52
	WM	SPANo	0.76	0.17	0.17	1.00	-1.02	0.76	0.98
		SPANr	0.66	0.22	0.01	1.00	-0.94	0.43	0.98
	SM	SMg	0.56	0.17	0.08	0.92	-0.05	-0.34	0.88
		SMp	0.62	0.26	0.00	0.97	-0.69	-0.66	0.96

Note. DRM = Deese–Roediger–McDermott paradigm; MC = Memory Conjunction paradigm; w = word stimuli; s = sentence stimuli; WM = working memory measures; SM = source monitoring measures; SPANo = operation span; SPANr = reading span; SMg = gender source memory; SMp = picture source memory, α = reliability estimate.

Analysis section in Method for details). Task level correlations and scatter plots of the primary dependent variables can be found in the Supplemental Material, and participant level data can be found on OSF.

Prior to reporting the primary analyses, for completeness we examined $\Delta d'$ across all task conditions (see Method for details). False memory sensitivity measures can be found in Table 2. Mean $\Delta d'$ was submitted to a linear mixed effect model with fixed effects of *paradigm* (DRM vs. MC), *stimulus type* (word vs. sentence), and *condition* (warning vs. no warning), as well as a random effect of participant. Discrimination of new items from lures was better in the DRM paradigm (*Paradigm*: $b = .55$, $SE = .05$, $p < .001$), with sentence stimuli (*Stimulus Type*: $b = .19$, $SE = .05$, $p < .001$), and without warnings (*Stimulus Type*: $b = .12$, $SE = .08$, $p = .045$). No other main effects or interactions were significant ($ps > .05$). These findings indicate clear differences in discriminability across paradigms and stimulus types, regardless of warnings.

Section 1: Do stable individual differences in false memories hold across different paradigms? To determine whether the rank ordering of performance within individuals changes across tasks or if there are stable individual differences in false memory, we specified two confirmatory factor

Table 2 Sensitivity measures for each false-memory task

Task	No Warning			Warning		
	New	Lure	$\Delta d'$	New	Lure	$\Delta d'$
DRMw	1.62 (0.05)	0.60 (0.04)	1.03 (0.05)	1.65 (0.06)	0.64 (0.05)	1.01 (0.06)
DRMs	1.05 (0.04)	-0.18 (0.03)	1.22 (0.04)	1.05 (0.04)	-0.05 (0.04)	1.10 (0.05)
MCw	2.44 (0.06)	1.93 (0.06)	0.51 (0.04)	2.35 (0.07)	1.92 (0.06)	0.43 (0.04)
MCs	1.58 (0.06)	0.90 (0.05)	0.67 (0.04)	1.65 (0.07)	0.97 (0.05)	0.68 (0.04)

Note. DRM = Deese–Roediger–McDermott paradigm; MC = Memory Conjunction paradigm; w = word stimuli; s = sentence stimuli.

models for each condition in which false memory, working memory, and source monitoring tasks loaded onto separate factors. The primary difference between the two models was the structure of the false-memory tasks. In the three-factor model, all false-memory tasks loaded onto a single false memory latent variable. This is a general false memory model, as it tests the hypothesis that the processes underlying false memory are largely invariant across task types. In the four-factor model, DRM (DRMw and DRMs) and MC (MCw and MCs) tasks loaded onto two separate factors. This is a task-specific model, as it tests the hypothesis that false memory differs as a function of task. In the factor analytic approach, the χ^2 statistic reflects how well the specified model reproduces the variance-covariance structure of the observed data. A significant χ^2 test ($p < .05$) is undesirable because it means that the theoretical model does not accurately reflect the observed structure. Other measures of goodness-of-fit indices are also reported. CFI and NFI values greater than .90, and SRMR and RMSEA values less than .08, are indicative of acceptable fit (Kline, 2015). As can be seen in Table 3, both models provide an acceptable fit to the data with and without warnings.

To determine the best fitting model, a series of χ^2 differences tests (notated as $\Delta \chi^2$) are performed. In the case of comparable fits among different theoretical models (i.e., no difference between $\Delta \chi^2$), it is customary to select the more parsimonious model (i.e., the simpler model with fewer factors). In the current study, there was no difference in model fit between the three-factor and four-factor models in either condition (*No Warning*: $\Delta \chi^2(3) = 1.81$, $p = .613$; *Warning*: $\Delta \chi^2(3) = 5.49$, $p = .139$). This indicates that specifying task-specific factors does not significantly improve model fit. The more parsimonious, and thus favored, general structure suggests there are stable individual differences in false remembering across tasks. As can be seen in Fig. 1,[2] working

memory and source monitoring are positively correlated with one another and higher performance on both tasks is associated with better discriminability in the false-memory tasks.

Finally, we tested for measurement invariance to ensure that the loadings onto the factors were roughly equal across warning conditions. To do so, we set "condition" (warning vs. no warning) as a grouping factor for the best fitting (three-factor) model and compared it to a model that had an additional constraint that the factor loadings be equal across conditions. Doing so did not significantly affect model fit, $\Delta \chi^2(5) = 8.82$, $p = .116$. This suggests that warnings did not change the factor structure.

Section 2: Does source monitoring mediate the relation between working memory and false memory? To determine whether the correlations between working memory and source monitoring with the general false-memory factor were shared or unique, we tested a mediation model where working memory predicted source monitoring, working memory predicted false memories, and source monitoring predicted false memories. As can be seen in Fig. 2, the relation between working memory and false memory was fully mediated by source-monitoring ability in both conditions (*No Warnings*: indirect effect = .23, $p = .001$; *Warnings*: indirect effect = .24, $p = .001$). Replicating Unsworth and Brewer (2010), these findings suggest that individuals with higher working memory did better on the false-memory tasks because they were better able to monitor for the origin of a memory.

Section 3: Is the source monitoring advantage associated with more effective lure rejection strategies? To determine whether the better discriminability for high source-monitoring ability participants was driven by qualitatively different lure rejection strategies, we submitted mean rates of rejection strategy (i.e., "recall-to-reject"; "would have remembered"; or "lacks familiarity") to separate linear mixed-effect models, with fixed effects of source-monitoring ability and condition and a random effect of subject (see Fig. 3). We collapsed across task because previous analyses indicated all false-memory tasks loaded onto a single factor. Disqualifying "recall-to-reject" strategies were used more by high ability participants (*Source Monitoring*: $b = .04$, $SE = .02$, $p = .013$),

---

[2] As can be seen in Fig. 1, the factor loading for the MCw task was not significant in either condition. Importantly, excluding this variable from the three-factor model did not significantly change the pattern of results. We therefore elected to keep all variables in the model for all analyses.

Table 3 Model fits for confirmatory factor analysis across conditions

Condition	Model	df	χ²	p	CFI	NNFI	SRMR	RMSEA [90% CI]
No Warning	**3 Factor**	**17**	**18.64**	**0.350**	**>.99**	**0.99**	**0.03**	**.02 [.00, .06]**
	4 Factor	14	16.83	0.265	0.99	0.98	0.03	.03 [.00, .07]
Warning	**3 Factor**	**17**	**23.39**	**0.137**	**0.98**	**0.96**	**0.04**	**.04 [.00, .08]**
	4 Factor	14	17.89	0.212	0.99	0.97	0.04	.04 [.00, .08]

Note. Model in boldface is best fitting model. df = degrees of freedom; χ² = chi squared; p = p value; CFI = comparative fit index; NNFI = nonnormed fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error.

whereas diagnostic "would-have-remembered" strategies tended to be utilized more by low-ability participants (*Source Monitoring*: $b = -.03$, $SE = .02$, $p = .102$). Usage of the "lacks familiarity" strategy was not associated with source-monitoring ability (*Source Monitoring*: $b = -.01$, $SE = .02$, $p = .522$). The only other significant effect was that disqualifying monitoring was more often used following warnings (*Condition*: $b = .04$, $SE = .02$, $p = .035$; all other main effects and interactions ps > .05). These findings indicate that high ability participants were able to recall information from study lists to determine that critical lures were not members of that list.

General discussion

The current study leveraged experimental and individual differences methodology to determine if a general framework could account for false recognition across two seemingly different word list paradigms, specify the processes underpinning false recognition in these tasks, and understand the strategies that are employed to avoid memory errors. There was clear evidence for consistency in false recognition across tasks, independent of warnings, that was predicted by both working-memory capacity and source-monitoring abilities. However, working-memory capacity was no longer associated with false recognition after controlling for source monitoring. The superior performance for high source-monitoring ability participants appears to be driven by the efficient usage of disqualifying monitoring processes. Below, we discuss the theoretical and applied ramifications of these findings.

The first aim of the current study was to examine whether two widely used false-memory paradigms were similar in their induction of memory errors. The results found that, at the task level, the ability to discriminate new items from critical lures was greater in the DRM compared with the MC paradigm. The two most prominent accounts of false memories in these paradigms (Brainerd & Reyna, 2001; Gallo, 2010) suggest that errors are jointly determined by processes at encoding (lure activation or gist extraction) and retrieval (monitoring or recollection rejection). The semantic/associative nature of the DRM task may have made it more likely that the lure was

activated, or a gist trace was formed, during encoding, making DRM lures more accessible. Alternatively, phonological familiarity may have decayed more rapidly than semantic familiarity (Matzen et al., 2011), making MC lures less compelling over time. It is therefore not entirely surprising that we found differences in discriminability across tasks. Critically, despite substantial differences at the *task* level, performance within an individual was relatively stable across tasks. That is, those who were more likely to falsely recognize lures in one task type were also more likely to falsely recognize lures in other tasks.

The finding that there were stable individual differences in false recognition across tasks is important for theory development and the debate as to whether individuals who are prone to false memories are so regardless of the paradigm used. In the current study and our previous work, false memories from multiple tasks all loaded onto single false-memory factor. In the current study, these were recognition errors occurring in different false-memory (DRM and MC) paradigms with different encoding contexts (words and sentences). In the Unsworth and Brewer (2010) study, these were recall errors not only from a DRM task but also from standard list-learning paradigms (free recall and paired associates) with related words, unrelated words, and numbers. The results from both studies also indicated that individuals higher in source-monitoring ability had better discriminability regardless of the types of details weighing into the decision (e.g., semantic vs. phonological; internal source vs. external source). These findings suggest that the commonality across tasks is due, at least in part, to a general memory-monitoring mechanism and that this monitoring mechanism is a stable individual difference. Importantly, however, we caution against extrapolating this interpretation beyond the word-list-learning paradigms used across the two studies.

As described previously, not all studies have found correlations between different task types (Falzarano & Siedlecki, 2019; Monds et al., 2017; Nichols & Loftus, 2019; Ost et al., 2013; Patihis et al., 2018; Salthouse & Siedlecki, 2007; Wilkinson & Hyman, 1998). In particular, when there are departures from list-learning paradigms, such as with text-based misinformation paradigms,
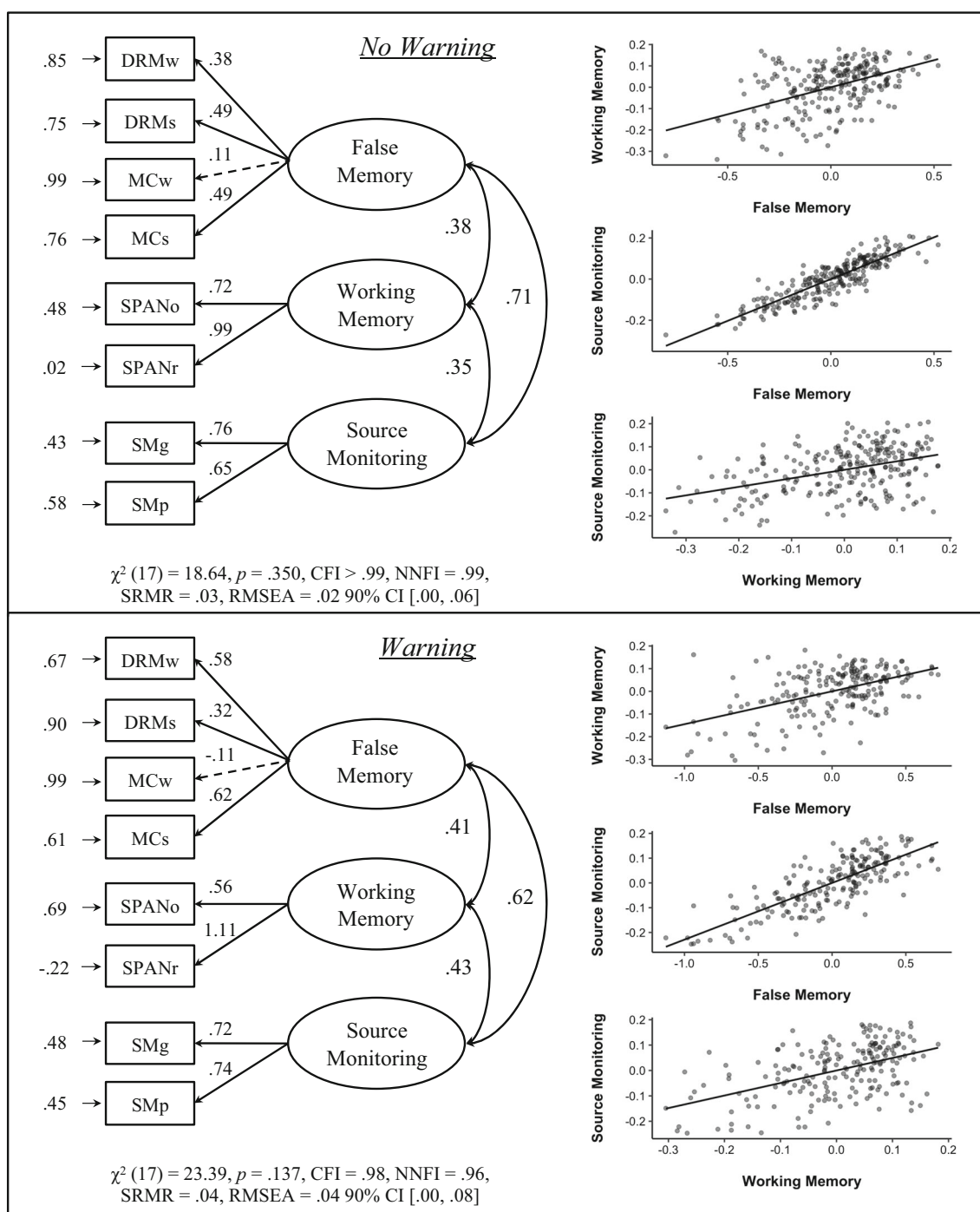
Fig. 1 Confirmatory factor analysis of the best-fitting three-factor model (left) and scatter plots of latent correlations (right) without (upper) and with (lower) warnings. Solid lines in factor analysis indicate significant paths at $p < .05$

these correlations are considerably weaker or nonexistent. This is somewhat surprising because prior research has indicated that source monitoring is an important process underlying many different false-memory paradigms, including suggestibility and misinformation paradigms (Lindsay & Johnson, 1989; Mitchell et al., 2003; Moore & Lampinen, 2016). It may be the case that a common

mechanism operates differently across in some paradigms (e.g., DRM and MC) than others (e.g., misinformation; Patihis et al., 2018). Another possibility is that prior research using only a single indicator of false memory or not accounting for potential response bias differences across tasks may preclude reliably finding correlations. These conflicting findings highlight the need for more
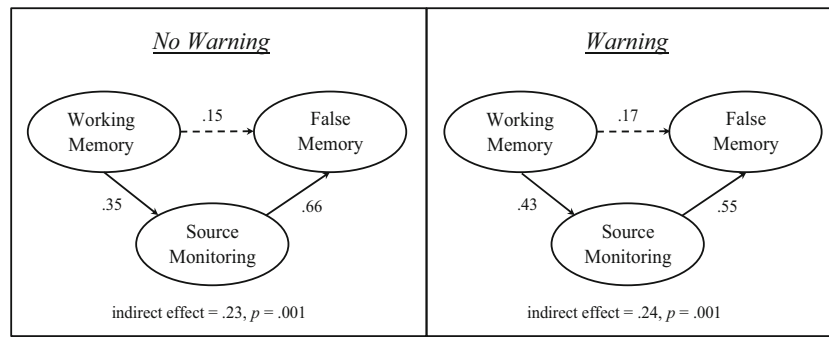
Fig. 2 Mediation analysis without (left) and with (right) warnings of false memory (Δd '), working memory, and source-monitoring ability. Solid lines in factor analysis indicate significant paths at $p < .05$, whereas dashed lines reflect non-significant paths. Source monitoring fully mediated the relation between working memory and false memory

research to better understand why sometimes these associations are found and sometimes they are not.

The current study also provides insights into the previously observed relations between working memory and false memory. Although the typical interpretation is that individuals with higher working memory are better able to monitor retrieval (Leding, 2012; Unsworth & Brewer, 2010), Watson et al. (2005) instead suggest that they are better able to maintain goals (i.e., "ignore lures"). This is based on the fact that in their study, working memory was only predictive of performance when participants were warned about the illusion during encoding. Our results are inconsistent with the attention
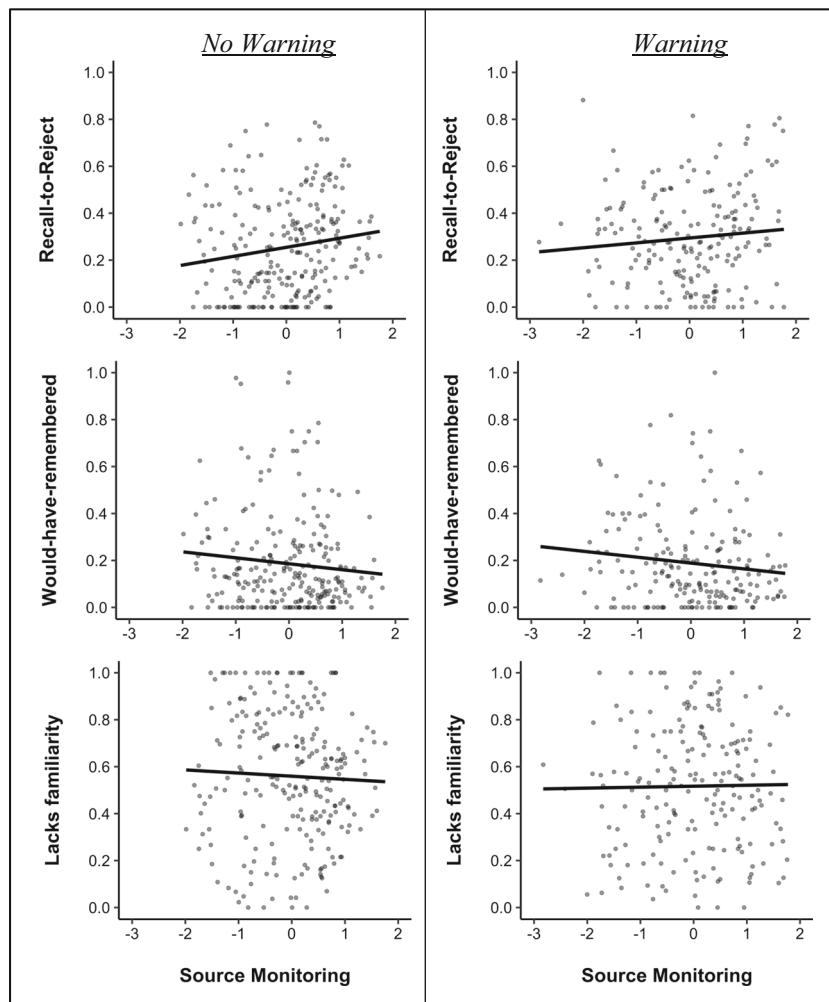

Fig. 3 Rejection strategies following critical lure rejection as a function of source-monitoring ability (z-scored) in each condition

account, as warnings did not change the relation between working memory and false recognition, and this relation was fully mediated by source monitoring. It should be noted, however, that warnings had little influence on discriminability at the task level. Prior literature on the efficacy of warnings generally shows that warnings prior to encoding reduce errors (Gallo et al., 1997; Gallo et al., 2001; McDermott & Roediger, 1998). It may be that the inclusion of rejection strategies at test (i.e., monitoring) along with performing multiple tasks (i.e., practice) reduced the efficacy of warnings to some degree. Future research should assess both attention and source-monitoring abilities in the same group of participants and use different types of warnings at encoding or retrieval.

Phenomenological reports following correct rejections of critical lures further suggest that variability in source-monitoring ability was associated with differing quality of retrieved details. In the present data, high source-monitoring-ability participants were more likely to use a disqualifying monitoring strategy, whereas those lower in source-monitoring ability tended to use a diagnostic monitoring strategy. These results are generally consistent with those from Leding (2012) that showed that high-ability participants were more likely to rely on diagnostic monitoring. Presumably, individuals with higher source-monitoring ability are better able to retrieve studied items or associated contextual features to reject critical lures. Low-ability participants may have greater difficulty in retrieving this contextual information. In cases in which retrieval fails, they may tend to rely more on heuristic decisions based on beliefs or plausibility. The current study is consistent with previous research indicating that disqualifying monitoring is critical for reducing false memories (Gallo, 2004), regardless of task type.

It is interesting to note that the gender (male/female) and picture (location) source monitoring tasks involved discriminating between two or more external sources. This is quite different from the false-memory tasks in which participants had to discriminate between external and internal sources of information stored in memory to avoid false remembering. That is, to reject critical lures, participants had to decide whether the item was perceived (e.g., *bed; tailspin*), merely imagined (e.g., *sleep*), or contained some but not all elements from study (e.g., *tailgate*; Johnson et al., 1993). Despite differences in the source dimension information (external-external vs. internal-external) contributing to these decisions, the correlation between source monitoring and false memory was quite high. What remains unclear is whether the source-monitoring tasks used in this study rely more heavily on diagnostic versus disqualifying monitoring. An interesting avenue for future research would be to examine rejection strategies in the source-monitoring tasks to determine whether individuals who are more likely to rely on disqualifying (or diagnostic) monitoring strategies are also more likely to use disqualifying (or diagnostic) monitoring strategies in the false-memory

tasks. Assuming monitoring is a stable individual differences factor, such relations should be found.

Conclusions

The results from the current study suggest that the ability to retrieve contextual information from a previous learning episode is integral to rejecting false memories in these two paradigms. Although we examined two commonly used and robust paradigms for eliciting false memories, research suggests that similar monitoring mechanisms may be critical in other domains, including eyewitness suggestibility (Lindsay & Johnson, 1989), misinformation (Mitchell et al., 2003; Moore & Lampinen, 2016), unconscious plagiarism (Marsh et al., 1997), and social contagion (Meade & Roediger, 2002). In each of these instances, source confusions (e.g., did I see the perpetrator or was he described to me?) may be more likely if contextual information (e.g., facial features) cannot be retrieved during recall. Monitoring deficits with increased age (Hashtroudi et al., 1989; Henkel et al., 1998; Pierce et al., 2008) may also contribute to greater susceptibility to false remembering (Gallo et al., 2006) and fake news (Brashier & Schacter, 2020). Thus, future work aimed at improving recollection-based monitoring processes may go far in reducing errors across a variety of important domains.

References

Alakbarova, D., Hicks, J. L., & Ball, B. H. (2021). The influence of semantic context on false memories. *Memory & Cognition*, 1–13. Advance online publication. https://doi.org/10.3758/s13421-021-01182-1

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior*, 22(3), 261–295.

Arndt, J., & Jones, T. C. (2008). Elaborative processing and conjunction errors in recognition memory. *Memory & Cognition*, 36(5), 899-912.

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, 48(1), 1–15.

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(4), 941–947. https://doi.org/10.1037/0278-7393.27.4.941.

Brainerd, C. J., & Reyna, V. F. (2001). *Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience.* In H. W. Reese & R. Kail (Eds.), Advances in child development and behavior (p. 41–100). Academic Press.

Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review, 24*(4), 396–439.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *27*(2), 307–327. https://doi.org/10.1037/0278-7393.27.2.307

Brashier, N. M., & Schacter, D. L. (2020). Aging in an era of fake news. *Current Directions in Psychological Science, 29*(3), 316–323. https://doi.org/10.1177/0963721420915872

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22.

Dodson, C. S., & Schacter, D. L. (2001). If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*(1), 155–161.

Falzarano, F., & Siedlecki, K. L. (2019). Investigating the relations among different measures of false memory. *Advances in Cognitive Psychology, 15*(4), 290–300. https://doi.org/10.5709/acp-0276-0

Gallo, D. A. (2004). Using Recall to Reduce False Recognition: Diagnostic and Disqualifying Monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 120–128. https://doi.org/10.1037/0278-7393.30.1.120

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks.* Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*(7), 833–848.

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14*(6), 730–741.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review, 4*(2), 271–276.

Gallo, D. A., & Roediger, H. L., III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*(3), 469–497.

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review, 8*(3), 579–586.

Gerrie, M. P., & Garry, M. (2007). Individual differences in working memory capacity affect false memories for missing aspects of events. *Memory, 15*(5), 561–571.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4*(1), 106–112.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments, & Computers, 27*(1), 46–51.

Hege, A. C. G., & 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*(4), 787–795. https://doi.org/10.1037/0278-7393.30.4.787

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127*(3), 251–268. https://doi.org/10.1037/0096-3445.127.3.251

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*(1), 1–18.

Holden, L. R., Conway, A. R., & Goodwin, K. A. (2020). *How individual differences in working memory and source monitoring matter in susceptibility to false memory* (Unpublished manuscript).

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541.

Jaschinski, U., & Wentura, D. (2002). Misleading postevent information and working memory capacity: An individual differences approach to eyewitness memory. *Applied Cognitive Psychology: The Official*

Journal of the Society for Applied Research in Memory and Cognition, 16(2), 223–231.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*(1), 82–102. https://doi.org/10.1006/jmla.2000.2761

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*(1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

Kline, R. B. (2015). *Principles and practice of structural equation modeling.* Guilford.

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 332–342. https://doi.org/10.1037/0278-7393.30.2.332

Leding, J. K. (2012). Working memory predicts the rejection of false memories. *Memory, 20*(3), 217–223.

Leding, J. K., & Lampinen, J. M. (2009). Memory conjunction errors: The effects of presentation duration and study repetition. *Memory, 17*(5), 597–607.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17*(3), 349–358.

Lloyd, M. E. (2007). Metamemorial influences in recognition memory: Pictorial encoding reduces conjunction errors. *Memory & Cognition, 35*(5), 1067–1073.

Lloyd, M. E. (2013). Reducing the familiarity of conjunction lures with pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1609–1614. https://doi.org/10.1037/a0031144

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7*(4), 560–572.

Loftus, E. F. (2019). Eyewitness testimony. *Applied Cognitive Psychology, 33*(4), 498–503.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25*(12), 720–725. https://doi.org/10.3928/0048-5713-19951201-07

Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*(2), 268–283.

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 886–897. https://doi.org/10.1037/0278-7393.23.4.886

Matzen, L. E., & Benjamin, A. S. (2009). Remembering words not presented in sentences: How study context changes patterns of false memories. *Memory & Cognition, 37*(1), 52–64.

Matzen, L. E., Taylor, E. G., & Benjamin, A. S. (2011). Contributions of familiarity and recollection rejection to recognition: Evidence from the time course of false recognition for semantic and memory conjunction lures. *Memory, 19*(1), 1–16.

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition, 30*(7), 1065–1077.

McDermott, K. B., & Roediger, H. L., III. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39*(3), 508–520.

Meade, M. L., & Roediger, H. L. (2002). Explorations in the social contagion of memory. *Memory & Cognition, 30*(7), 995–1009.

Meyersburg, C. A., Bogdan, R., Gallo, D. A., & McNally, R. J. (2009). False memory propensity in people reporting recovered memories of past lives. *Journal of Abnormal Psychology, 118*(2), 399–404. https://doi.org/10.1037/a0015371

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). Oxford University Press.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*(4), 638–677. https://doi.org/10.1037/a0015849

Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 17*(1), 107–119.

Monds, L. A., Paterson, H. M., & Kemp, R. I. (2017). Do emotional stimuli enhance or impede recall relative to neutral stimuli? An investigation of two "false memory" tasks. *Memory, 25*(8), 945–952.

Moore, K. N., & Lampinen, J. M. (2016). The use of recollection rejection in the misinformation paradigm. *Applied Cognitive Psychology, 30*(6), 992–1004.

Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermottfalse-memory paradigm: The role of identifiability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 35–41. https://doi.org/10.1037/0278-7393.29.1.35

Nichols, R. M., & Loftus, E. F. (2019). Who is susceptible in three false-memory tasks? *Memory, 27*(7), 962–984.

Odegard, T. N., & Lampinen, J. M. (2005). Recollection rejection: Gist cuing of verbatim memory. *Memory & Cognition, 33*(8), 1422–1430.

Odegard, T. N., Lampinen, J. M., & Toglia, M. P. (2005). Meaning's moderating effect on recollection rejection. *Journal of Memory and Language, 53*(3), 416–429.

Ost, J., Blank, H., Davies, J., Jones, G., Lambert, K., & Salmon, K. (2013). False memory ≠ false memory: DRM errors are unrelated to the misinformation effect. *PLOS ONE, 8*(4), Article e57939.

Parker, S., Garry, M., Engle, R. W., Harper, D. N., & Clifasefi, S. L. (2008). Psychotropic placebos reduce the misinformation effect by increasing monitoring at test. *Memory, 16*(4), 410–419.

Patihis, L., Frenda, S. J., & Loftus, E. F. (2018). False-memory tasks do not reliably predict other false memories. *Psychology of Consciousness: Theory, Research, and Practice, 5*(2), 140–160. https://doi.org/10.1037/cns0000147

Peters, M. J., Jelicic, M., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *European Journal of Cognitive Psychology, 19*(2), 213–232.

Pierce, B. H., Waring, J. D., Schacter, D. L., & Budson, A. E. (2008). Effects of distinctive encoding on source-based false recognition: further examination of recall-to-reject processes in aging and Alzheimer disease. *Cognitive and Behavioral Neurology: Official Journal of the Society for Behavioral and Cognitive Neurology, 21*(3), 179–186.

Platt, R. D., Lacey, S. C., Iobst, A. D., & Finkelman, D. (1998). Absorption, dissociation, and fantasy-proneness as predictors of memory distortion in autobiographical and laboratory-generated memories. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 12*(7), S77–S89.

Pope, K. S. (1996). Memory, abuse, and science: Questioning claims about the false memory syndrome epidemic. *American Psychologist, 51*(9), 957–974. https://doi.org/10.1037/0003-066X.51.9.957

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*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Salthouse, T. A., & Siedlecki, K. L. (2007). An individual difference analysis of false recognition. *The American Journal of Psychology, 120*(3), 429–458.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5*(4), 710–715.

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*(1), 71–86.

Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62*(1), 19–34.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505.

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 76–85.

West, E., & Meterko, V. (2015). Innocence project: DNA exonerations, 1989–2014: Review of data and findings from the first 25 years. *Albany Law Review, 79,* 717.

Wilkinson, C., & Hyman, I. E., Jr. (1998). Individual differences related to two types of memory errors: Word lists may not generalize to autobiographical memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 12*(7), S29–S46.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., & Dong, Q. (2013). The relationship between DRM and misinformation false memories. *Memory & Cognition, 41*(6), 832–838.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., Li, H., Xue, G., Lu, Z., & Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory, 18*(5), 543–555. https://doi.org/10.1080/09658211.2010.487051