Faculty and Student Publications   Psychology

5-1-2020

When do pictures reduce false memory?

Rebekah E. Smith
University of Mississippi

R. Reed Hunt
University of Mississippi

Recommended Citation

Smith, R. E., & Hunt, R. R. (2020). When do pictures reduce false memory? Memory & Cognition, 48(4), 623–644. https://doi.org/10.3758/s13421-019-00995-5

When do pictures reduce false memory?

Rebekah E. Smith[1] · R. Reed Hunt[1]

Abstract

An important discovery in false-memory research is Israel and Schacter's (*Psychonomic Bulletin & Review, 4*, 577-581, 1997) finding that presenting pictures at study relative to words alone reduces false memory in the DRM paradigm, a result that has been replicated many times. The standard interpretation is that memory for visual processing of the pictures can be used to reject the critical distractors, which were not explicitly present at study. Beginning from the empirical observation that the pictures used by Israel and Schacter are not consistently labelled with the DRM word they are supposed to represent, we present a series of four studies designed to determine if it is the presentation of pictures or the mismatch between the pictures and the words that reduces false memory. The results across the four experiments demonstrate that picture presentation at study is neither necessary nor sufficient to reduce false memory in the DRM and the categorical associate paradigms. However, we discuss other studies in which picture processing clearly is responsible for reduction of false alarms and note that these studies use study materials and memory tests that are different from the DRM and categorical associate paradigms in that critical lures are externally provided rather than generated. We speculate that the effectiveness of memory for visual processing for reducing false memory may depend on the source of the false memory, but this remains for future research.

Keywords False memory · DRM · Pictures · Modality

Introduction

Roediger and McDermott's (1995) resurrection of Deese's (1959) associated word list paradigm (DRM) initiated a flurry of research activity on false memory. The now familiar laboratory preparation entails presentation of words for study that are all related to the same critical word, which itself is not presented. Both recognition and recall tests for the studied items reliably yield false memory for the critical non-presented item. In addition to providing a simple, reliable technique for producing false memory, the DRM paradigm has face validity as a good model for the effect of inference on memory, in that the critical non-presented item comes to mind during presentation of the study list, similar to an inference coming to mind in comprehension. The challenge for memory is discriminating presented material from self-generated material.

Research to date has produced a considerable body of useful information about this form of false memory. Not only is the non-presented critical item frequently remembered, in some cases at the same rate as studied items, confidence in that false memory tends to be high (Lampinen, Neuschatz, & Payne, 1997; Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger & McDermott, 1995). False memory for the critical item is relatively long lasting (Payne et al., 1996; Seamon, Luo, Kopecky, Price, Rothschild, Fung, & Schwartz, 2002; Thapar & McDermott, 2001). Warnings about the possible false memory have little effect unless given prior to the study list (Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001b; McCabe & Smith, 2002; McDermott & Roediger, 1998; Neuschatz, Benoit, & Payne, 2003), and when a false memory occurs, corrective feedback does little to reduce future occurrence (McConnell & Hunt, 2007; Mullet & Marsh, 2016).

In the face of this resilience, some variables can significantly reduce false memory for the critical DRM item (e.g., Arndt & Reder, 2003; Israel & Schacter, 1997; Smith & Hunt, 1998). One of the most commonly studied of these variables is the presentation at study of pictures accompanying the auditory presentation of the DRM lists. Compared to auditory presentation alone, false memory is markedly reduced by picture presentation (Israel & Schacter, 1997). Subsequent to Israel and Schacter's discovery, the beneficial effect of pictures has

✉ Rebekah E. Smith
  rebekah@olemiss.edu

[1] Department of Psychology, The University of Mississippi, Oxford, MS 38677, USA

been replicated numerous times with the DRM paradigm (Dodson & Hege, 2005; Ghetti, Qin, & Goodman, 2002; Hege & Dodson, 2004; Hicks & Starnes, 2006; Howe, Cicchetti, Toth, & Ceritto, 2004; Howe, 2008; Schacter, Cendan, Dodson, & Clifford, 2001; Schacter, Israel, & Racine, 1999; Smith, Hunt, & Dunlap, 2015). All of these studies used Israel and Schacter's (1997) picture-word pairs. Those picture-word pairs have been generously shared with other laboratories, and the data resulting from their use have been quite reliable.

The importance of the picture effect itself is as the discovery of one of the first variables to reduce false memory, and an important function of the follow-up studies was to demonstrate the reliability of the effect. A second important impact of the follow-up studies was to develop (Schacter et al., 1999) and refine the distinctiveness heuristic, and more generally to draw attention to the theoretical importance of output monitoring in false memory. That attention has led in some cases to suggested modifications of the original theory (e.g., Hege & Dodson, 2004; Howe, 2008; Smith, Hunt, & Dunlap, 2015). For our current purposes, the critical point is that all of the studies used the same Israel and Schacter (1997) materials, and all assumed the same fundamental basis for the reduction in false memory. The generally accepted interpretation of these data is that the visual sensory processing required for pictures at study is remembered and then used at test to discriminate studied items from non-presented items. A monitoring process, the distinctiveness heuristic (Schacter et al., 1999), examines test items for memory of prior visual sensory processing. In the absence of such evidence, the test item can be rejected and because critical lures in false memory paradigms have not been visually processed at study, the prior visual processing required by pictures effectively reduces false memory.

Converging lines of evidence for the importance of visual processing include demonstrations that visual presentation of study list words leads to fewer false memories in the DRM paradigm than auditory presentation of study words (Cleary & Greene, 2002; Gallo, McDermott, Percer, & Roediger, 2001a; Gallo & Roediger, 2003; Hunt, Smith, & Dunlap, 2011; Kellogg, 2001; Smith & Engle, 2011; Smith & Hunt, 1998; Smith, Hunt, & Gallagher, 2008; Smith, Lozito, & Bayen, 2005). The reduction in DRM false alarms following visual presentation of the study words is consistent with the notion that memory for visual sensory processing can reduce false alarms relative to auditory processing.

Another line of evidence is studies showing that imagery instructions reduce false memory to critical items in DRM lists relative to auditory presentation (e.g., Foley, Wozniack, & Gillum, 2006; Foley et al., 2009).[1] Analogous to the picture

effect, the effect of imagery has been interpreted through the distinctiveness heuristic, where at test studied words are remembered to have been imagined, allowing rejection of test items whose representation contains no evidence of prior imagination. Foley et al. (2009) note that the information used to monitor test items may be the representation of an image or it may be memory for cognitive processes that produced the image, a distinction that parallels the situation with pictures where the useful information could be memory of a picture or memory for processes yielding comprehension of the picture.

Although much evidence exists for the beneficial effect of visual processing in the DRM paradigm, a definitive conclusion is complicated by studies that have found no effect of picture presentation on false alarms. One such report comes from the DRM paradigm (Howe, 2008) and two others from the closely related categorical associate paradigm (Koutstaal & Schacter, 1997; Koutstaal, Schacter, & Brenner, 2001). A potentially important difference between these three studies and the myriad of studies replicating Israel and Schacter's picture effect is that studies conducted by Howe and Koutstaal and colleagues did not use the Israel and Schacter line drawings as pictures. Rather the studies failing to report reduction in false alarms used actual pictures depicting the referents of the studied words. In contrast, the Israel and Schacter pictures seem to vary wildly in the fidelity with which they depict the referent of the DRM word. This is not surprising given the variability in the concreteness of the DRM words, but the important point is that we are unaware of data on the normative relationship between the DRM list words and their corresponding Israel and Schacter pictures. The need for such normative data can be illustrated by examining a subset of the pictures. For example, Fig. 1 shows a subset of the pictures used to represent study items from the critical associate SWEET. Items in the top row obviously represent Candy, Cake, Pie, respectively, but we suspect that items on the bottom row are much less obvious to people who have not spent considerable time with the materials. The issue of the match between the picture label and the DRM study word is important because if a large proportion of the pictures do not evoke the label represented by the DRM word, the interpretation of the picture effect as the result of visual sensory processing is open to question.

We gathered normative data on 18 of the 21 associative groupings used by Israel and Schacter (1977) by asking participants to label the pictures used to represent the DRM words in these categories. The methodological details of this study and the proportion of participants giving the DRM label for each picture can be found in Appendix A. Less than a third of the pictures were labeled with the DRM word by 80% or more of participants, a similar proportion of pictures were given the DRM word as label by less than 20% of participants, and over 10% of pictures were not given the DRM word as a label by a single participant. The overall average percentage of people

---

[1] Generating an image of each item individually produced a greater reduction in false memory than did generating integrated imagery involving multiple study items in the image (Foley et al., 2006, 2009).
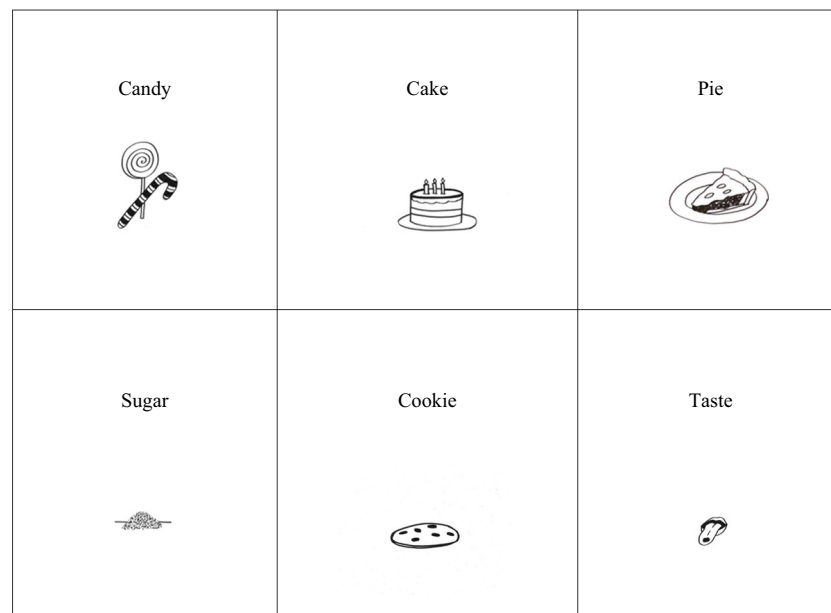
Fig. 1 Examples of pictures used by Israel and Schacter (1997). The associated DRM study-list words are shown above each picture. Pictures are available online from the Schacter Memory Lab https://scholar.Harvard.edu/schacterlab/pages/publications

labelling a picture with the corresponding DRM word was 44%, when including all of the items in Appendix A. The norming study demonstrates that many of the Israel and Schacter (1997) pictures may not be spontaneously perceived as the referent of the DRM word.[2] Thus, the door is open to a test of the necessity of visual processing of the picture to obtain the widely reported results from using the Israel and Schacter's pictures in the DRM paradigm.

Experiment 1

The normative data establish that most of the Israel and Schacter pictures may not spontaneously evoke their corresponding DRM word as a label. If those pictures spontaneously evoke a label that is different from the simultaneous auditory presentation of the DRM word, a confound is introduced in the comparison of the picture condition with a control condition that sees and hears exactly the same word. The picture condition now differs from the control not only because of the picture presentation but also because of a mismatch between the word and pictures. Thus, we do not know whether it is the picture

itself or the possible mismatch between pictures and words that leads to false memory reductions. The following experiments investigate the role of pictures and mismatching of study list stimuli presented in reducing false memory. The principle variables are type of study materials, words versus pictures, and the normative match between the auditory target word and the visually presented material, be it word or picture. The test is to determine what combination of materials produces the same outcome as found in studies using the Israel and Schacter pictures, i.e., reduced false alarms to critical distractors and no reduction in hit rate relative to a control condition.

The first experiment compared both recall and recognition of DRM lists for four different conditions (see overview in Table 1). In all conditions participants heard the DRM list word during study. In the auditory only control condition participants heard the words only. In the matched word condition, participants heard the DRM word and also saw the same word visually (e.g., heard "cake" and saw the word "cake"). In the picture condition, participants were also presented with the Israel and Schacter (1997) pictures while hearing the DRM words. In the fourth condition, the normative word condition, the most frequently generated normative label for the picture was presented with the auditory presentation of the DRM study list words.

The auditory control and picture conditions match the conditions normally used in much of the past research. We expected to replicate the results of previous research showing a reduction in false memory following picture presentation and no effect on correct memory for study list items. The normative word condition is designed to mimic the psychological state created if Israel/Schacter drawings recruit a label different from the intended DRM word but critically, in the absence

---

[2] As described in Appendix A, we elected to present the items in random order in the norming study, rather than blocked by theme as in Israel and Schacter (1997), because our goal was to identify items that were mismatches with the DRM words, excluding gist-based guesses. A reviewer noted that the random ordering of the items in the norming study could underestimate correct identification in a blocked presentation, thus, we refer to the possibility that the pictures may not be spontaneously identified as the DRM list item. Nonetheless, given the considerable differences between the normative label and the DRM list word in some cases, it seems very likely that the label will not be spontaneously generated in many cases even with blocked presentation.

Table 1 Overview of design and summary of recognition results for the current experiments

	Study conditions (participants in all conditions heard the study word list)	N	Reduced relative to auditory-only control?	
			False alarms	Hits
Exp. 1 Associatively Related DRM List	Auditory Only	55	-	-
	Matched Word	51	No	No
	Picture [a]	48	Yes	No
	Normative Word [b]	51	Yes	Yes
Exp. 2 Categorized List	Auditory Only	39	-	
	Matched Word	39	No	No, higher
	Matched Picture	40	No	No, higher
	Related Picture	41	Yes	Yes
			Reduced relative to matched visual word?	
			False alarms	Hits
Exp. 3 Categorized List	Matched Word	50	-	-
	Matched Picture	50	No	No
	Related Word	50	Yes	Yes
	Related Picture	51	Yes	Yes
Exp. 4 Categorized List	Matched Word	40	-	-
	Matched Picture	41	No	No
	Related Picture	40	Yes	Yes
	Half Matched/Half Related Picture	40	Yes	No

[a] The pictures in Experiment 1 were the Israel and Schacter (1997) pictures

[b] The label for each of the Israel and Schacter (1997) pictures that was most frequently produced by participants in the norming study (see Appendix A and Appendix B)

of visual processing of the picture. The matched word condition also requires visual processing of the DRM word, but is designed to eliminate the potential conflict between the picture and label. Thus, if sensory processing of pictures responsible for reduction of false memory in the DRM paradigm, only the picture condition should show the effect. If, however, visual processing of any material is sufficient to reduce false memory, both the normative word and matched word conditions should produce the same effect as the picture condition. Finally, if the effect of the Israel/Schacter pictures is due to a mismatch between the DRM word and a conflicting label evoked by the line drawing, the normative word condition will yield the same results as the picture condition.

Method

Participants and design GPower 3 (Faul, Erdfelder, Lang, & Buchner, 2007) was used to established targeted sample sizes of between 148 and 200 participants in each experiment to provide power of .90 to .97 to detect effects similar to those in recent related studies (e.g., Smith, Hunt, & Dunlap, 2015). The participants in all experiments were undergraduates who volunteered for course credit. The number of participants in each condition

for all experiments can be found in Table 1, along with an overview of the experimental conditions in each experiment.

The 205 participants in the first experiment were randomly distributed across the four conditions in roughly equal number. The study list words were presented auditorally in all conditions. In the auditory only condition, the words were not accompanied by visual stimuli. In the picture condition the Israel and Schacter (1997) pictures were presented with each auditory word (e.g., participants heard "taste" and saw the associated picture shown in Fig. 1). In the normative word condition, the most frequently produced label from our norming study was presented with the auditory DRM word (e.g., participants heard "taste" and saw the word "tongue"). Finally, in the matched word condition, the DRM study list word was also presented visually (e.g., participants heard "taste" and saw the word "taste").

Materials The study list contained six of the DRM categories in Israel and Schacter's (1997) materials. The categories Sweet, Sleep, Fruit, Mountain, Needle, and Chair were selected because of their relatively high rate of false memory production and the lack of overlap of words across the studied items in each list. The first 12 items were selected for each category, resulting in a single 72-item study list. The

normative words were the most frequently given label in the norms mentioned earlier and of the 72 DRM study words, 32 did not match the norm word. Mismatching pairs appeared in all six of the DRM categories with the number of mismatching pairs ranging from three to eight per category. The DRM words and their normative pair are shown in Appendix B.

The recognition test contained 42 items. Two study items from each DRM category (studied items) and all six (critical lures) of the unpresented critical items were on the test. In addition, the two most frequent DRM items from each category that were not present at study (weakly related lures) were included along with 12 items selected from DRM lists not used at study (unrelated lures). The test items for each associative list were presented in a blocked fashion with each block beginning with the first study list item, followed by a weakly related lure, an unrelated lure, a second study list item, a second unrelated lure, a second weakly related lure, and the critical item. The blocked organization was not obvious to the participants as all items were presented as a single list. This recognition test procedure matches that used in prior studies (e.g., Roediger & McDermott, 1995; Smith & Hunt, 1998).

All experiments in this study were programed with E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002).

To support possible replication efforts, the materials and instructions for the experiments in this paper can be found in the appendices. Data will be made available upon request, consistent with policies of the Psychonomic Society and the American Psychological Association.

Procedure After signing consent forms, each participant was instructed to put on a head set and listen to a series of digits for the purpose of adjusting the volume to their satisfaction. After adjusting the volume, participants saw instructions for the study lists on the monitor. The auditory only control was instructed that they would hear a list of words that they should try to remember. The matched word condition was instructed that they would hear a list of words and simultaneously see each word on the monitor. The picture condition was instructed that they would hear a list of words and simultaneously see a related picture. The normative word condition was instructed that they would hear a list of words and simultaneously see a related word. In each of the latter three conditions, the instructions were to pay attention to both components of the item for a later test. These were the same instructions used by Israel and Schacter (1997).

The visual items were shown for 1.5 s and accompanied by the recorded auditory presentation of the DRM study word read in a female voice. Following presentation of the single 72-item study list, participants were given a sheet of paper and asked to recall as much of the studied materials as possible. The recognition test was administered following the recall test. Test items were shown on the monitor and subjects were instructed to press the Y key if the item had been presented at

study and the N key if not. The test was self-paced. Instructions for all experiments can be found in Appendix C.

Results

Critical false recall The recall data are presented first. The mean proportions of study list items correctly recalled and DRM critical items falsely recalled are shown in the top portion of Fig. 2 as a function of study condition. False recall of the critical items clearly differed across the study conditions, $F(3,201) = 15.956$, $MSE = .023$, $p < .001$, $\eta_p^2 = .192$. Fig. 2 shows a pattern of decreasing false recall from auditory only to matched visual word to pictures to normative word. Multiple comparisons with Tukey's HSD test generally were consistent with the visual pattern, although the statistical outcome was slightly more complicated. The picture and normative word conditions led to reliably lower false recall than the auditory only condition (both ps < .001). The matched word condition did not differ reliably from the auditory only ($p = .115$) or the picture condition ($p = .178$), but reliably more false recall occurred in the matched word condition than in the normative word condition ($p < .001$). Finally, the difference in false recall between the picture and normative word conditions was not reliable ($p = .131$).

Correct recall For the auidtory only and matched word conditions the possible number of study list items to be recalled is 72. Given our norming study, one might argue that there are more possible items to be considered as part of the study episode for the picture and normative word conditions, however, we adopt the conventional method of scoring and consider 72 to also be the maximum number of items to be recalled correctly in the picture and normative word conditions. To match all conditions on what was scored as correct recall, we scored correct recall as the proportion of auditory items recalled for each participant, consistent with prior studies using the DRM pictures. These data are shown on the right side of the top portion of Fig. 2. Although the normative word condition had numerically reduced correct recall, the main effect of study condition did not reach significance, $F(3, 201) = 2.441$, $MSE = .003$, $p = .065$, $\eta_p^2 = .035$.

Critical false recognition The mean proportions of DRM critical items incorrectly receiving a Yes response on the recognition test is shown in the lower portion of Fig. 2. Study condition clearly affected performance, $F(3, 201) = 11.52$, $MSE = .062$, $p < .001$, $\eta_p^2 = .147$. Comparisons among the conditions using Tukey's HSD revealed that the auditory only condition produced more critical false alarms than the picture ($p = .001$) and the normative word ($p = .003$) conditions. Likewise, the normative word condition also led to more critical false alarms than the picture ($p < .001$) and the normative word ($p < .001$) conditions. The normative word and
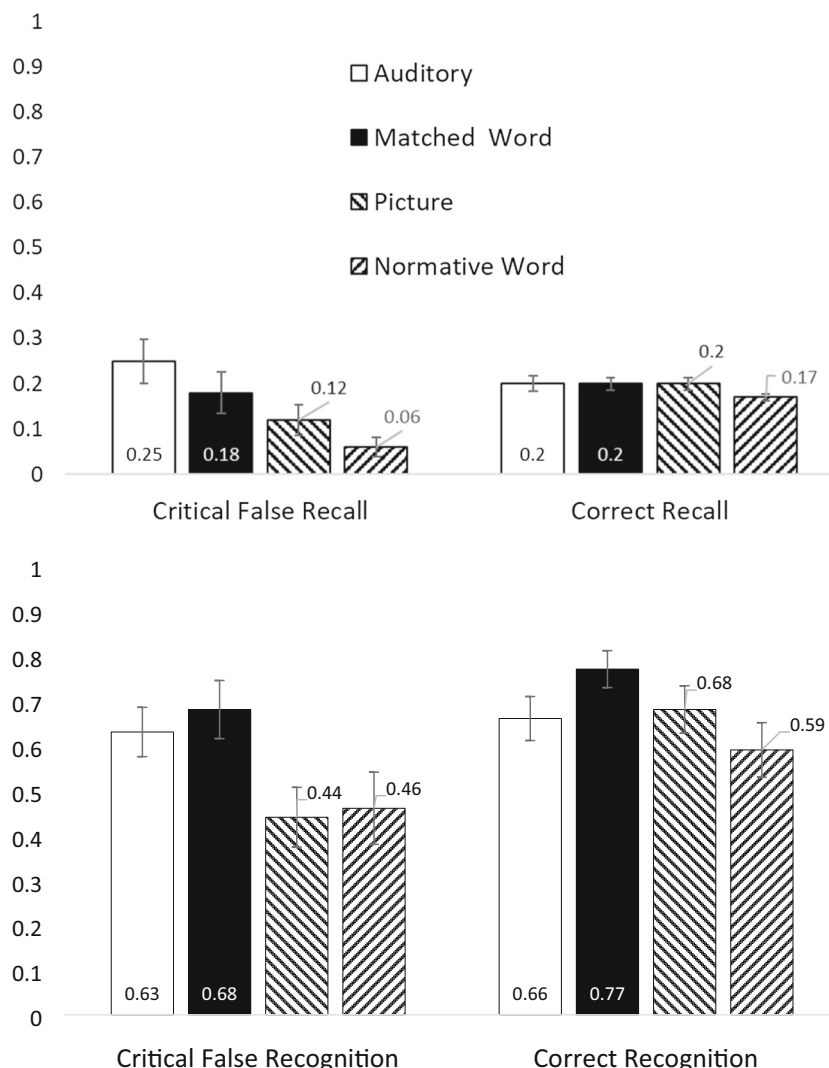
Fig. 2 Mean proportion of study-list items correctly recalled and DRM critical items falsely recalled (top) and correct and false recognition (bottom) in Experiment 1. Pictures are from Israel and Schacter (1997) and normative words are labels provided to these pictures in norming study described in text. Error bars represent 95% confidence intervals

picture conditions did not differ from one another ($p = .967$) nor did the auditory only and matched word conditions ($p = .809$).

Related and unrelated false alarms The mean proportions of false alarms to related and unrelated lures can be found in Table 2. The proportion of false alarms to related items did not differ reliably among the study conditions, $F(3, 201) = 2.44$, $MSE = .014$, $p = .066$, $\eta_p^2 = .035$. Study condition did affect unrelated item false alarms, $F(3, 201) = 2.90$, $MSE = .006$, $p = .036$, $\eta_p^2 = .041$. Subsequent Tukey comparisons showed that the only reliable effect was fewer unrelated false alarms in the picture condition than in the auditory only condition ($p = .026$; all other $ps > .16$).

Correct recognition The proportion of correct responses to study items can also be seen in the lower portion of Fig. 2.

The difference between study conditions was reliable, $F(3, 201) = 7.06$, $MSE = .038$, $p < .001$, $\eta_p^2 = .095$. Tukey comparisons among the conditions indicated that the matched word condition produced more hits than auditory only ($p =$

Table 2 Proportion of false alarms to related and unrelated lures on the recognition test in Experiment 1

Study conditions	Lure type			
	Related		Unrelated	
	M	SD	M	SD
Auditory Only	.19	.02	.06	.01
Matched Word	.18	.02	.05	.01
Israel & Schacter (1997) Pictures	.18	.02	.02	.01
Normed Word	.23	.02	.05	.01

.025) and normative word ($p < .001$) conditions. Hit rates in the picture condition did not differ reliably from any of the other three conditions nor did hits differ in the auditory only and normative word conditions, ps > .10.

Discussion

Picture presentation at study effectively reduced false memory for critical distractors in both recall and recognition, replicating previous findings using the DRM paradigm (Dodson & Hege, 2005; Ghetti, et al., 2002; Hege & Dodson, 2004; Hicks & Starnes, 2006; Howe et al., 2004; Howe, 2008; Israel & Schacter, 1997; Schacter et al., 2001; Schacter et al., 1999; Smith et al., 2015). Importantly, at the same time, presentation of the normative word also resulted in a reduction in false memory. Consequently, it is tempting to conclude that the reduction in false memory in the picture condition and the normative word condition is due to the mismatch between some of the visual and auditory pairs. Such a conclusion would suggest that pictures are not necessary for false memory reduction and point to other theoretical interpretations. The reasoning is that the same mismatches occur between the auditory-visual pairs in the picture and normative word conditions and hence the same mechanism causes the reduction in false memory in the two conditions.

However, such a conclusion here is premature. Perhaps the sensory processing of the pictures exerts an effect over and above the effect of the mismatches, or the participants in the picture condition may not even perceive the mismatches that are explicit in the normative word condition. Thus, although the data are consistent with the hypothesis that sensory processing of pictures is not necessary for reduction of false memory in the DRM paradigm, they do not compel this conclusion. As it turns out, DRM lists are not well suited to asking about the role of pictures in reducing DRM false memories, a point addressed in our remaining experiments.

Experiment 2

For many of the DRM words there is no direct veridical picture of the concept referenced by the word. Thus, picture presentation of DRM words inevitably confounds the effect of pictures with the effect of a mismatch between some number of the DRM words and their corresponding pictures. This problem can be avoided by using categorized lists of words and their pictures for which picture-label match norms exist (e.g., Szekely et al. 2003; Szekely et al., 2004). From these norms, enough words can be selected for which pictures are well matched to form study lists. Many prior studies have demonstrated powerful false memory effects following study of categorized lists where memory for non-presented category instances was the index of false memory (e.g., Brainerd &

Reyna, 1998; Dewhurst, 2001; Howe, 2006; Kouststaal, Schacter, & Brenner, 2001; Smith, Ward, et al., 2000; Smith, Gerkins, Pierce, & Choi, 2002).

The second experiment used categorized lists for which pictures have been normed. These materials allow a direct comparison between the effect of pictures that evoke a label matching the exemplar word versus pictures whose labels do not match the exemplar, and thereby allow a direct contrast between the effect of picture presentation with the effect of a mismatch between picture and word. All participants heard a list of category instances. In the first condition, the auditory only condition, participants heard all of the categorized list words at study with no accompanying visual presentation. In the second condition, the matched word condition, participants heard the words and simultaneously saw the same word on the monitor. There were two conditions in which participants heard the words presented with pictures. In the matched picture condition, auditory presentation of the word was accompanied by simultaneous visual presentation of a picture of the object referenced by the word. In the related picture condition, the auditory presentation of the word was accompanied by a picture of an object related to the word. For example, the word *horse* was accompanied by a picture of a horse in the matched picture condition and by a picture of a saddle in the related picture condition. If visual sensory processing pictures is responsible for reduction of false memory, the matched picture condition should yield fewer false alarms than the auditory only and matched word condition. Alternatively, if the mismatches between the labels for studied pictures and words is responsible, only the related picture condition will produce fewer false alarms. The matched word condition is included as a control for the effect of visual processing of a word.

Predictions for the hit rates in the auditory only, the matched picture and word conditions are straightforward. Each item appears twice for the latter two conditions and only once for the auditory only condition, and therefore we expect higher hit rates in the matched picture and word conditions than in the auditory only condition. Predictions for the related picture condition are less obvious. Two different items are presented on each trial, and that difference could detract from study item processing, leading to relatively poor performance. Alternatively, participants may attempt to reconcile the discrepancy between the items on each trial, and the processing required for that reconciliation may enhance correct recognition.

Method

Subjects and design The subjects, 158 volunteers from Introductory Psychology, were randomly assigned to one of four experimental conditions defined by the study list presentation. All conditions heard the study words. The matched

picture condition saw pictures of objects referenced by the study word. The related picture condition saw a picture related to the study word (see Appendix D). The matched word condition saw the study list word, and the auditory only condition only heard the words. The recognition memory test consisted of study list words, related distractors from the same category as the study list and unrelated distractors from different categories.

Materials The 64 study list words were drawn from category norms (Battig & Montague, 1969; Van Overschelde, Rawson, & Dunlosky, 2004) with eight words taken from eight different categories. The selection of the words was constrained by the availability of a picture of the object referenced by the word in the International Picture Naming Project (IPNP) database (Szekely et al. 2003; Szekely et al., 2004). The first eight words after the most frequent example of the category that met the picture match criterion were selected for study. The criterion for picture-name match was at least 80% of the norming subjects named the picture with the study list word.

The related pictures were objects obviously related to the study words (e.g., horse-saddle, eye-eyeglasses), whose picture-name agreement with their own label also was set at a minimum of 80%. Potential related items were selected by a research assistant and then reviewed by two other research assistants and the lab director independently. Items were discarded and replaced until the all four were in agreement that there was a relationship between the categorized item and the potential related item.

Study-list presentation was blocked by category. The complete list of study words and the names of the related pictures is presented in Appendix D, along with the picture name agreement for each category picture and its name and each related picture and its name. The picture-name agreement information in Appendix D comes from the International Picture Naming Project (IPNP) database (Szekely et al., 2003, 2004) and references the naming agreement between the labels and the picture.

The test items were 128 words presented visually. All 64 of the study items plus 64 distractors comprised the test list. The critical distractors were 32 instances from the studied categories not presented at study. These items were the four most frequent instances from each studied category not used in the study list. The unrelated distractors were eight instances taken from each of four non-studied categories. The 32 unrelated items were selected so as not to overlap with studied words or pictures. The lure items can be found in Appendix E.

Procedure Following the consent process and sound check, participants were given instructions for the study phase consistent with their study condition (see Appendix C).

Participants were presented with the study list items one at a time for 2 s each.[3] Words were blocked by category and presented in declining order of frequency. Prior to the recognition test, participants completed a 3-min filler task in which letters were shown one at a time in the middle of the screen and participants pressed the key for the next letter in the alphabet as quickly as possible.[4] Recognition test words were shown in the middle of the screen and participants pressed the Y key to indicate an item had been presented during the study phase and press the N key to indicate that it had not been present at study. Following a response, the next item appeared. Test items were shown in a random order.

Results

Critical false recognition Consistent with the visual depiction of critical false alarms in Fig. 3, study condition reliably affected performance, $F(3, 154) = 6.33$, $MSE = .035$, $p < .001$, $\eta_p^2 = .110$. Tukey's comparisons showed the related picture condition produced fewer critical false alarms than the auditory only ($p = .003$) and the matched word ($p = .001$) conditions, and a trend towards lower critical false alarms in the related picture condition relative to the matched picture condition ($p = .072$). No other individual comparisons were lower than $p = .45$.

False alarms to unrelated items The mean proportion of false alarms to unrelated items was quite low ($M = .05$, $SE = .01$), and the effect of study condition on performance was not reliable, $F(3, 154) = 1.04$, $MSE = .006$, $p = .38$, $\eta_p^2 = .020$.

Correct recognition The hit rates in Fig. 3 show a difference in the proportion of study items correctly recognized as a function of study condition, and this difference was reliable, $F(3, 157) = 14.899$, MSE $= .018$, $p < .001$, $\eta_p^2 = .225$. Tukey's comparisons revealed that related pictures produced lower hit rates than the auditory only ($p = .037$), matched word ($p < .001$), and matched picture ($p < .001$) conditions. The hit rate

---

[3] Studies using the standard DRM paradigm employ a variety of study item presentation times. For Experiment 1, we selected 1.5 s to match procedures we had used successfully in prior studies (e.g., Smith & Hunt, 1998). Studies using categorized lists to investigate false memories also employ a variety of presentation times, including self-paced encoding tasks (Smith et al., 2002), 1 s (e.g., Dewhurst, 2002), and 3 s (e.g., Smith et al. 2000). Importantly, Koutstaal et al. (2001) conducted a study with categorized lists and picture stimuli and found that the pattern of results remained the same across presentation durations ranging from 1.5 s to 2 s. We opted to use a 2-s rate in Experiments 2 through 4 as this appeared to be a mid-point between prior studies using timed study presentation rates with young adult participants.

[4] The filler task was used to avoid ceiling effects in correct memory. Although we selected to include a recall test in Experiment 1 to parallel our earlier work on the modality effect in the standard DRM paradigm (Smith & Hunt, 1998), we chose not to use a recall text in Experiments 2 through 4 because prior research has also shown that the presence of an intervening recall test can influence the pattern of modality effects on subsequent recognition tests (Smith, Hunt, & Gallagher, 2008).
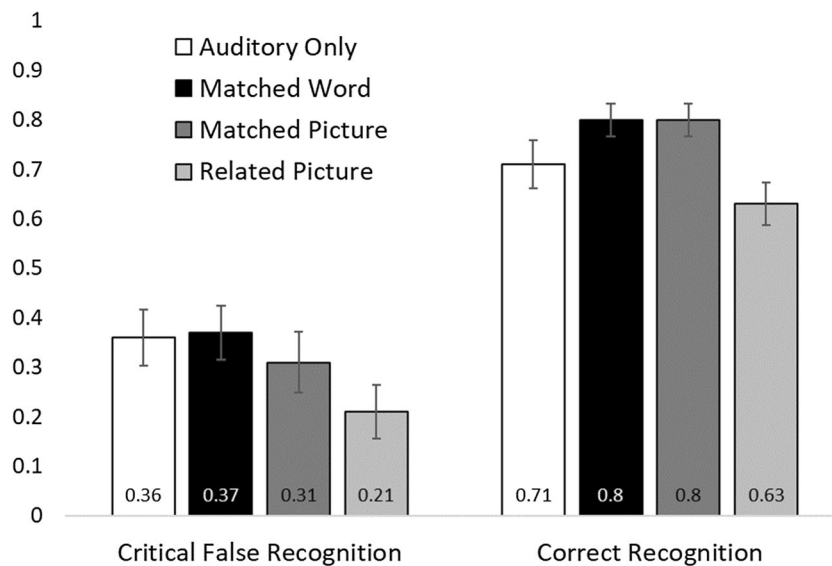
Fig. 3 Mean proportion of correct and false recognition in Experiment 2. Error bars represent 95% confidence intervals

in the auditory only condition was reliably lower than in the matched word ($p = .027$) and matched picture ($p = .015$) conditions. The matched word and matched picture conditions did not differ on correct recognition of study of items, ($p = .997$).[5]

Discussion

The related picture condition produced the lowest rate of false alarms. In contrast, the matched picture condition, where the pictures matched the word, did not differ from the auditory only control on related false alarms. Hit rates followed the opposite pattern with poorest performance in the related picture condition. Hit rates for matched word and matched picture conditions were equal and quite high. This result is not surprising given the redundancy of the visual and auditory items.

The failure to find any reduction in critical false alarms in the matched picture condition is surprising in the context of the dominant interpretation of the picture effect. Importantly, this result replicates other research from both the DRM paradigm (Howe, 2008) and the categorical associate paradigm (Koutstaal & Schacter, 1997; Koutstaal et al., 2001).

---

[5] A reviewer noted that participants may have confused some lure items with pictures that they had seen at study. For instance, the related lure "toe" may have been called old because participants had seen a picture of the study list item "foot," which may have elevated the false alarm rate in the conditions shown the matched pictures. To address this possibility, we reviewed the study item pictures and lure words and identified the following items that may have resulted in some confusion between studied items and lure items: foot, toe, screwdriver, chisel, jacket, cost, shirt, ax, hammer, wrench, pliers, grapefruit, and grape. After excluding these items and recalculating hit rates and false alarm rates, we reanalyzed the results for Experiments 2, 3, and 4. In all cases, the pattern of results fully matched the pattern of results reported in the manuscript.

Common to all of these experiments is that they did not use the Israel/Schacter pictures but rather used pictures whose labels more closely matched the auditorally presented word. One would not expect this outcome if memory for visual sensory processing of pictures was sufficient to reduce false alarms.

Alternatively, the substantial reduction in false alarms in the related picture condition is consistent with the hypothesis that the picture effect on false memory using Israel/Schacter drawings is driven by the mismatch between the spontaneously evoked label for the drawing and the auditorally presented DRM word. However, the current experiment does not address the possibility that the reduction in false memory in the related picture condition is still a function of seeing a picture at study. This is addressed in the next experiment by including a condition in which the auditory word is accompanied by the word that is the label of the related picture. If the presentation of the related information, regardless of whether a picture is presented, is the essential factor in reducing false memories, then both the related picture and related word conditions should show reductions in false memory.

Experiment 3

The third experiment replicates Experiment 2 with a slight change in the design. The auditory only control condition was dropped. We added a new related word condition in which auditory presentation of the target word was accompanied by visual presentation of the related word. This related word condition allows us to determine if the reduction in false alarms in the related picture condition is limited to picture presentation of the related item.

Method

Subjects and design A total of 201 student volunteers were distributed across four conditions of study. All conditions heard the study list words. In the matched word and matched picture conditions the visually presented item matched the auditorally presented word. In the other two conditions, related pictures and related word, the visually presented item was related to, but not the same as, the auditorally presented word.

Materials and procedure The materials were the same as those used in Experiment 2. The label for the related picture was the word used in therelated word condition. The procedure was the same as in Experiment 2.

Results

Critical false recognition The results are summarized in Fig. 4, which shows the mean proportion of critical false alarms along with the hits as a function of study condition. As depicted in Fig. 4, study condition affected the proportion of critical false alarms, $F(3,197) = 8.31$, $MSE = .025$, p < .001, $\eta_p^2 = .112$. Individual comparisons via Tukey's HSD showed that matched picture and matched word conditions did not differ ($p = .96$) nor did the related picture and related word conditions differ ($p = .74$). Related picture study led to fewer related false alarms than both the matched picture ($p < .001$) and the matched word ($p = .002$) conditions. Likewise, related word study yielded fewer related false alarms than matched picture ($p = .01$) and matched word ($p = .04$) conditions.

**Unrelated false alarms** As in Experiment 2, the rate of unrelated false alarms was quite low ($M = .06$, $SE = .004$), and the rate of unrelated false alarms did not differ across conditions, $F(3, 197) = 1.13$, $p = .337$.

Correct recall The pattern of hits in Fig. 4 mirrors that of the related false alarms in that the conditions favorable for reduction of false alarms also reduce hits. The effect of study condition on proportion of hits was reliable, $F(3,197) = 14.22$, $MSE = .027$, $p < .001$, $\eta_p^2 = .178$. Tukey's HSD comparisons showed that hits did not differ between matched picture and matched word conditions ($p = .98$) nor between related picture and related word conditions ($p = .34$). The matched picture study condition led to more hits than therelated picture ($p = .001$) and the related word conditions ($p < .001$). The matched word condition also produced reliably more hits than the related picture ($p = .004$) and the related word conditions ($p < .001$).

Discussion

Taken together, the results of the second and third experiments show that false alarms are much lower when pictures are only related to studied words than when the pictures are identical to the words. Indeed, the matched picture condition in Experiment 2 yielded as many false alarms as the auditory only control condition. Moreover, as shown in Experiment 3, the pictures in the related picture condition are not necessary for the beneficial effect on false alarms because related words shown visually produce the same effect. In short, the pattern of results in these two experiments suggests that picture processing *per se* is neither necessary nor sufficient for reducing false alarms.

There remains, however, an important inconsistency between our data and those produced by studies using the Israel/Schacter pictures in DRM lists. In those previous
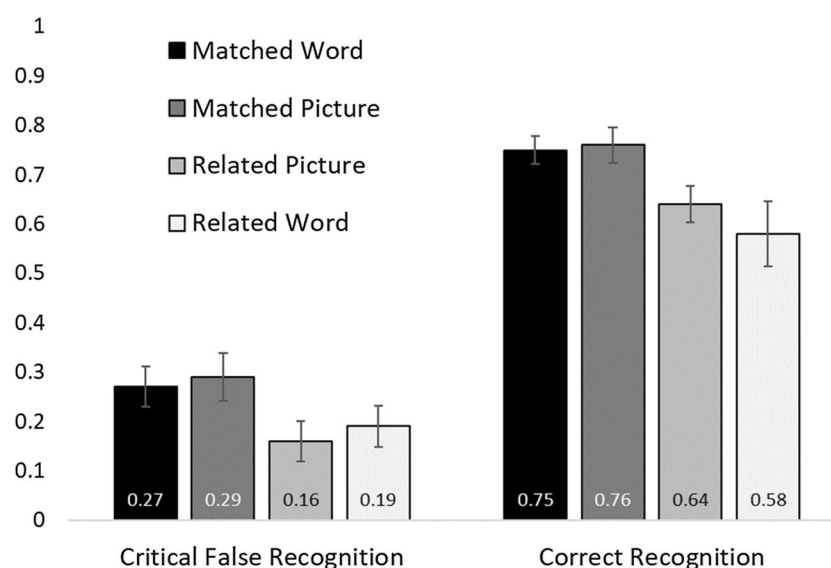
Fig. 4 Mean proportion of correct and false recognition in Experiment 3. Error bars represent 95% confidence intervals

studies, pictures reduced false alarms but had no effect on hit rates, a result we replicated in Experiment 1. In Experiments 2 and 3, the condition designed to model the mismatch between picture and label in the DRM list, related picture condition, produced significant reductions in correct recognition. The inconsistency between the effect of the related picture condition on categorized lists and the picture condition on DRM lists questions the validity of generalization from the former to the latter. The difference between results of previous studies using Israel/Schacter pictures and the related picture condition in Experiments 2 and 3 may well be due to the fact that some of the pictures in the Israel and Schacter set do match the DRM word, as demonstrated in our normative study of those pictures.

Experiment 4

The fourth experiment explores this possibility by creating a half-matched/half-related picture study list condition. In this condition the matched picture was shown for half of the items and the related picture was shown for the other half of the items, more closely modelling the relationship between DRM words and the Israel/Schacter pictures. Perhaps including matching pictures in the list encourages the attempt to reconcile the mismatch between the word and related picture, and the elaborative processing required for reconciliation will improve the hit rates relative to Experiments 2 and 3.

Method

Participants and design A total of 162 students from Introductory Psychology participated in return for credit toward a course assignment. The participants were distributed randomly in approximately equal number across four study conditions, all of which heard the categorized study list words. In the matched picture condition, the pictures all matched the auditory study words. In the half match/half related condition, half of the pictures matched the auditory study words and half were only related to the auditory study words. The proportion of matching pictures approximates the proportion of Israel/Schacter pictures that matched their normative labels for the DRM lists used in Experiment 1 (.52). In the related pictures condition, none of the pictures matched the auditory study words. Finally, in the matched word condition, the same word was presented auditorally and visually.

Materials and procedure The pictures and words were the same as those used in Experiments 2 and 3, and the lists for the matched picture, related picture, and matched word correspond to the lists in the previous experiments. The new condition, half-match picture, saw study lists consisting of one half matching and one half related pictures, which were

formed by using the matched pictures for half of each category's instances and related pictures for the other half. For the specific DRM words used in Experiment 1, the label provided in the norming study matched the DRM list word for the associated picture 51% of the time. Thus, the 50% match in the current experiment closely approximates the overall likely match rate for the DRM picture condition in Experiment 1. Two separate lists were created such that all studied words were heard in the presence of matched pictures by half of the participants and in the presence of related pictures by half of the participants. The auditory presentation lists were the same in all conditions, allowing comparison among the conditions for memory of the auditory words as a function of the accompanying visual presentation.

Otherwise the procedure, including the recognition test, was identical to that used in Experiment 3.

Results

Critical false recognition Figure 5 shows the proportion of critical false alarms as a function of study condition. Beginning with the analysis of critical false alarms, the effect of study condition depicted in Fig. 5 was reliable, $F(3, 158) = 11.07$, $MSE = .021$, $p < .001$, $\eta_p^2 = .174$. Individual comparisons via Tukey's HSD revealed that the half match condition led to fewer critical false alarms than both the matched picture ($p = .010$) and the matched word ($p < .001$) conditions. Likewise, the related pictures condition also produced fewer critical false alarms than both the matched picture ($p = .011$) and the matched word ($p < .001$) conditions. The half match and related picture conditions did not differ reliably ($p = 1.00$) nor did the matched picture and matched word conditions differ ($p = .407$).

Unrelated false alarms As in the prior two experiments, false alarms to unrelated items were infrequent ($M = .06$, $SE = .01$) and did not differ as a function of study condition $F(3, 158) = 1.80$, $MSE = .009$, $p = .149$.

Correct recognition The analysis of correct recognition of study list items, also shown in Fig. 5, yielded a reliable effect of study condition, $F(3, 158) = 10.84$, $MSE = .018$, $p < .001$, $\eta_p^2 = .171$. Individual comparisons using Tukey's HSD showed that the related picture condition produced the lowest level of hits compared to the other three conditions, with $p < .01$ for all three comparisons. The other three conditions did not differ on hits (all $ps > 23$).

For the half match condition, we also compared correct recognition for the matched items (e.g., hear *horse* and see picture of a *horse*; $M = .71$, $SE = .02$) and the related items (e.g., hear *horse* and see picture of a *saddle*; $M = .72$, $SE = .02$), and found no difference between these two item types, $t < 1$, $p > .57$.
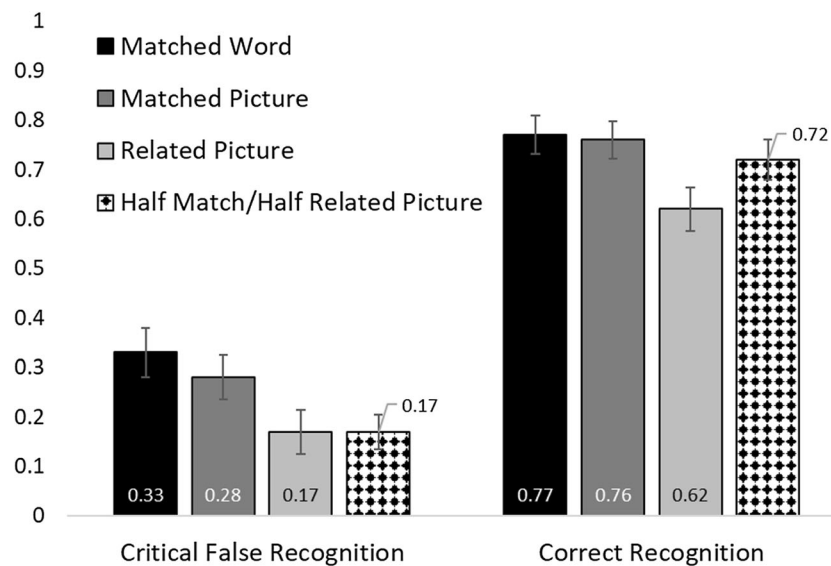
Fig. 5 Mean proportion of correct and false recognition in Experiment 4. Error bars represent 95% confidence intervals

Discussion

The results from both false alarms and hits for the matched picture, related picture, and matched word conditions replicated those of Experiments 2 and 3. Critical false alarms among these conditions were lowest when none of the pictures matched the auditory word, but so were the hits. The new data from Experiment 4 are those from the half match condition. Critical false alarm rates in this condition were equivalent to those of the condition where none of the pictures matched the word, the lowest rates among the conditions. Unlike the related picture condition, however, the hit rate for the half match was equivalent to the matched picture and matched word conditions.

The pattern of hits and false alarms produced in the half match condition is the same as the pattern reported in all previous studies using the Israel and Schacter (1997) pictures, critical false alarms were reduced relative to auditory only presentation of the study list, but hits were equivalent to the auditory only conditions. In contrast, the matched picture condition did not reduce critical false alarms compared to controls, replicating the results of the second and third experiments. This consistent finding suggests that visual processing of pictures is neither necessary nor sufficient to produce reductions in false alarms (again false alarms in the matched picture condition were equivalent to the matched word condition).

General discussion

The discovery that presenting pictures at study reduces false memory in the DRM paradigm (Israel & Schacter, 1997) is an important contribution to our understanding of false memory.

The experiments reported here were designed to test the assumption that visual sensory processing of pictures is responsible for the reduction. The investigation was motivated by the observation that many of the pictures used in previous DRM experiments are not spontaneously labelled with the corresponding DRM study word, thus introducing the confounding factor of picture-word incongruity for many of the studied pairs. In order to systematically separate the effects on false memory of picture presentation from picture-word incongruity, categorized study lists were substituted for DRM lists because the normative data indicated that too few of the Israel-Schacter drawings elicited the DRM word as a label, likely because many of the DRM words are abstract. In contrast, available norms provide plenty of pictures representing category instances whose spontaneously evoked label is the same as the word representing the instance, and previous research has shown that categorized lists provide a close analogy to the DRM paradigm for study of false memory.

Reducing false memories in the DRM paradigm: Pictures are neither necessary nor sufficient

Thus, armed with the ability to systematically manipulate the relationship between the label elicited by the picture and the presented study word, our experiments were designed to directly compare the effect of visual sensory processing of pictures with the effect of picture-word incongruity. The results clearly show that pictures are neither necessary, as indicated by the equivalent reduction in false alarms with all mismatching words, nor sufficient, as shown by the findings that showing all matching pictures did not reduce false alarms. Of note is the reliability of this null result, which was obtained in three different experiments. Thus, we suggest that effect of

the half match condition in Experiment 4 and by extension the effect of pictures in the DRM paradigm, is not caused by the sensory visual processing of the pictures, but rather is due to processes engaged by the mismatch between some of the pictures and corresponding list words.

It is important to note that the conclusion that pictures are not sufficient for reducing false memories appears to be limited to the DRM types of paradigm. We return to this point following our discussion of theoretical interpretations of our current findings.

Theoretical interpretations of the current findings

These experiments were not designed to address the theoretical issue of how the mismatch reduces false alarms, and the data cannot do so *post hoc* simply because the overall pattern, reduced false alarms and equivalent hits relative to control, is consistent with the two broad categories of mechanisms. The two categories are memory-based processes and decision-based processes (Arndt & Reder, 2003) or as Jacoby and associates characterize them, early selection versus late correction (Jacoby, Shimizu, Daniels, & Rhodes, 2005). The primary example of decision-based processing is the distinctiveness heuristic (Schacter et al., 1999), a heuristic process that monitors representations of test items for evidence of their presence at study. When information the individual considers memorable is encoded and the monitoring heuristic is engaged at test, each item is examined for that information, and if it is absent in the representation of the test item, that item is rejected.

Several variations of the memory-based account have been proposed, the most frequently mentioned of which explains the reduction of false alarms as the result of enhanced item-based processing of the study list. One version of this approach suggests that enhanced item-based processing detracts from relational processing among list items, effectively minimizing activation of critical lures during study (e.g., Arndt & Reder, 2003; Huff & Bodner, 2013). At odds with this idea are the results of numerous studies of item-specific and relational processing (Hunt & McDaniel, 1993). The only instance in which item-specific processing interfered with relational processing is when the relational dimension is defined as list serial position (McDaniel & Bugg, 2008). In contrast, numerous experiments that directly index relational processing along semantic dimensions show no reduction as the result of item-specific processing (e.g., Hunt & Einstein, 1981; Hunt & Seta, 1984). The only conditions in our experiments where relational processing likely is disrupted are the related picture and related word conditions of Experiments 2–4 where the visual and auditory items never match. Here both false alarms and hits are reduced but that is likely due to disruption of all processing, much like that found following a non-semantic orienting task (Hunt et al., 2011). Another variation of the

item-based, relational processing interpretation assumes that enhanced item-based processing provides information that is monitored at test to oppose the effect of relational processing as it pertains to the critical lures. This approach is essentially the same as the distinctive heuristic. Another possibility is enhanced processing at study, combining item and relational information, yields a highly diagnostic memory and with appropriate cues at test, access is limited to study items, an example of Jacoby's early selection principle (Gray & Gallo, 2015; Hunt et al., 2011). Finally, it should be noted that data from several studies indicate that manipulations reducing false memory affect both decision-based and memory-based processes (Hege & Dodson, 2004; Huff, Bodner, & Faucett, 2015; Hunt et al., 2011). But again, conclusions concerning the best interpretation of the mismatch effect await further research.

Pictures are sufficient for reducing some types of false memory

We note a potentially important caveat, namely that our data and conclusions are not consistent with all studies of pictures and false memory. While we are not the only ones who have reported no reduction in false alarms when pictures with consensual labels are presented at study (e.g., Koutstaal & Schacter, 1997; Koutstaal, Schacter, & Brenner, 2001, with categorized lists; Howe, 2008, with DRM lists), other studies have found substantial reduction in false alarms following study of pictures with consensual labels (Dodson & Schacter, 2002; Gallo, Cotel, Moore, & Schacter, 2007; Gallo, Foster, & Johnson, 2009; Gallo, Weiss, & Schacter, 2004; Scimeca, McDonough, & Gallo, 2011). What accounts for the differences in the two sets of findings?

Different effects as a function of the source of false memories

Although methodological differences are numerous across the studies, one likely possibility is the combination of the type of materials and the paradigm in which those materials are embedded. The studies reporting a reduction in false memory with veridical pictures used unrelated study materials and either a running recognition or a criterial recollection test. The critical false alarms in these two tests are semantically unrelated to study items but they are pre-exposed prior to the test. Specifically, lures were repeated in the running recognition test. In tests of criterial recollection, test cues define the properties of studied items that are to be recognized and critical lures are items that were present at study but that lack the criterial properties defined by the cue. These studies require distinguishing between different types of externally presented items. In contrast, studies reporting no reduction in critical false alarms following picture presentation at study have used

the categorical association and DRM paradigms, coupled with a standard yes/no recognition test. In these latter paradigms the critical lure has not been explicitly presented, but generally is assumed to have been implicitly elicited during study by the semantic relationship among study items (Roediger, Watson, McDermott, & Gallo, 2001). The activated information is semantically integrated with the externally supplied information, rendering discrimination between internally generated and externally presented material difficult.

Cast in source-monitoring terms, the discrimination in the association paradigms (the DRM and categorized lists used in the current experiments) is between external (studied) and internal (lures) items. Johnson, Hastroudi, and Lindsay (1993) argued that memory for complex cognitive operations is more important for distinguishing external-internal sources than for discriminating between two external sources. In addition, Johnson (2006) has argued that automatically activated inferences, such as those underlying the false memory generation in the DRM paradigm and the categorized list presentations in Experiments 2, 3, and 4 of the current experiment, do not involve complex cognitive operations. Indeed, Johnson, Raye, Foley, and Foley (1981) showed that participants are biased to attribute familiar items to external sources rather than to self-generation, suggesting that the familiarity aroused by a critical lure in the association paradigms is highly likely to be attributed to its presence in the study list.

Conclusion

The sum of these factors implies that the discrimination between externally presented studied items and internally generated lures is more difficult in the association paradigms than is distinguishing between two externally presented types of items in the running recognition and criterial recollection paradigms. Thus, pictures paired with highly consensual labels may fail to reduce false alarms in the association paradigms because minimal complexity is required for visual processing of pictures that are redundant with their accompanying labels. However, mismatched pictures do require complex processing in the form of semantic elaboration to reconcile the inconsistency between picture and label. That complex cognitive operation does facilitate discrimination of the perceptually processed study items and the internally generated lure. Although speculative, our explanation has testable implications as well as offering a plausible reason for the robustness of false memory for inferences.

None of the data or materials for the experiments reported here are available, and none of the experiments was preregistered. Data will be made available upon request.

Appendix A: Norming study method and results

We conducted a norming study to determine what label participants would apply to the pictures from Israel and Schacter (1997) when viewing the pictures alone without the associated DRM word.

Methods

Participants The 78 participants completed the study in exchange for credit towards a requirement in an Introductory Psychology course.

Materials and procedure We used 18 of the 21 associative categories from Israel and Schacter (1997). Fifteen of these categories have 12 pictures available, two categories had 11 pictures, and one had 13 pictures for a total of 215 pictures. The categories were divided into two sets of nine categories, with one set having 108 pictures and the other 107. The pictures in each set were randomized and half of the participants saw each set.

Participants, who completed the study in two groups, one with 38 participants and the other with 40 participants, were seated at desk in a classroom and were instructed that they would see pictures shown on the classroom screen and that their task was to write down a single word for each picture that identified what the picture depicted. Participants were told that there was no right or wrong answer. Each picture was shown for 15 s. The two groups were run in the same afternoon.

Results

We compared the responses for each picture with the DRM word for that picture. In some instances, the response and DRM word were the same. We also counted response that included the DRM word as a match. For instance, for the DRM word "ache," responses such as "aches," "backache," and "stomach ache" were included as a match. However, for many items very few of the responses matched the DRM list. For two dozen pictures, not a single participant labeled the picture with the DRM word and only one picture "hat" was labeled with the DRM by all participants. Fig. 6 shows the distribution of pictures with relatively poor to relatively good matches between the normed label responses and the DRM words. Table 3 shows the proportion of participants who responded with the DRM match word for each individual picture.
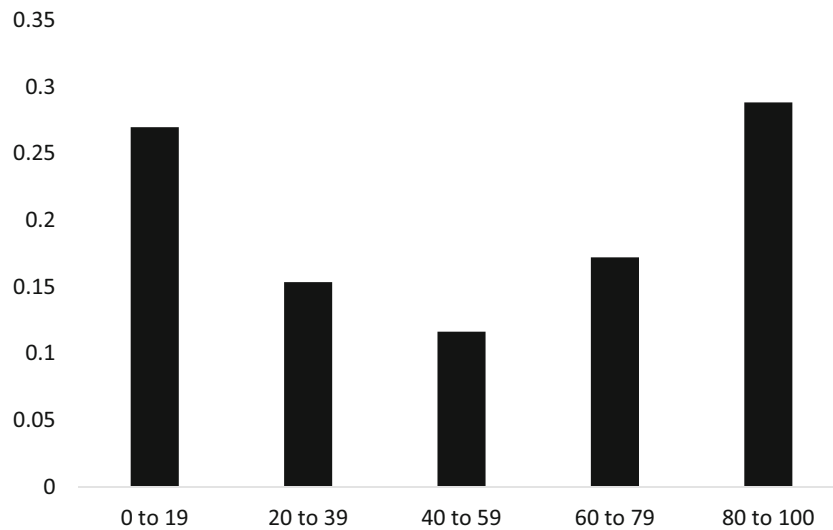
Fig. 6 Distribution of pictures with respect to match between norming study response and DRM word associated with the pictures. The left most bar represents the proportion of pictures (.27) labeled with the DRM word by less than 20% of participants. The right most bar represents the proportion of pictures (.29) with for which at least 80% of participants labeled the picture with the DRM word

Table 3 The proportion of participants providing a picture naming response that matched the DRM word associated with that picture. The mean proportion of participants (along with the standard deviations) are reported for each DRM category.

DRM Category and Study List Words		Proportion	DRM Category and Study List Words		Proportion
baby	bottle	0.92	Chair	bench	0.97
Mean = .54	boy	0.58	Mean = .48	couch	0.79
SD = .37	crib	0.21	SD = .37	cushion	0.00
	cry	0.74		desk	0.42
	diaper	0.05		legs	0.87
	doll	0.97		recliner	0.58
	girl	0.79		rocking	0.00
	mother	0.03		seat	0.37
	rattle	0.79		sit	0.00
	small	0.03		stool	0.68
	spoon	0.92		swivel	0.21
	toys	0.50		table	0.97
bath	bubble	0.88	doctor	bill	0.23
Mean = .42	clean	0.10	Mean = .42	dentist	0.85
SD = .31	dirty	0.28	SD = .25	hospital	0.23
	room	0.43		lawyer	0.08
	shower	0.85		medicine	0.38
	soap	0.43		nurse	0.65
	swim	0.78		office	0.40
	tub	0.13		patient	0.00
	wash	0.60		pills	0.45
	water	0.13		sick	0.64
	wet	0.05		stethoscope	0.55
bread	butter	0.80		surgeon	0.58
Mean = .36	crust	0.00	eagle	america	0.29
SD = .35	dough	0.25	Mean = .49	bald	0.08
	eat	0.23	SD = .37	beak	0.84
	flour	0.00		bird	0.13
	food	0.30		claw	0.61
	jam	0.28		coin	0.37
	loaf	0.03		flag	0.87
	milk	0.65		fly	0.00
	sandwich	0.90		nest	0.68
	slice	0.03		scout	0.89
	wine	0.88		sky	0.32
				wings	0.87

Table 3 (continued)

DRM Category and Study List Words		Proportion	DRM Category and Study List Words		Proportion
foot	arm	0.03	house	barn	0.84
Mean = .50	boot	0.80	Mean = .75	boat	0.92
SD =.37	hand	0.60	SD = .27	brick	0.66
	inch	0.35		building	0.39
	kick	0.10		car	0.92
	sandals	0.85		chimney	0.87
	shoe	0.88		door	0.95
	soccer	0.75		family	0.92
	sock	0.85		garage	0.95
	toe	0.35		key	0.84
	walk	0.70		street	0.03
	yard	0.00		windows	0.87
fruit	apple	0.90	moon	beam	0.00
Mean = .52	banana	0.90	Mean = .42	bright	0.00
SD = .34	basket	0.78	SD = .37	cheese	0.93
	berry	0.25		crater	0.15
	bowl	0.65		earth	0.55
	cherry	0.85		light	0.48
	cocktail	0.08		love	0.13
	kiwi	0.28		planet	0.23
	orange	0.40		rocket	0.83
	pear	0.88		shine	0.00
	salad	0.23		stars	0.83
	vegetable	0.03		sun	0.88
head	ache	0.21	mountain	bike	0.95
Mean = .67	brain	0.92	Mean = .56	cave	0.79
SD = .31	cap	0.61	SD = .36	cliff	0.66
	ear	0.21		climb	0.29
	hair	0.66		desert	0.00
	hat	1.00		goat	0.79
	lettuce	0.18		hill	0.82
	neck	0.79		magic	0.55
	nose	0.97		rain	0.82
	scarf	0.95		ski	0.92
	shoulders	0.63		stream	0.05
	toilet	0.92		valley	0.05
music	art	0.23	sleep	awake	0.30
Mean = .38	band	0.63	Mean .45	bed	0.83
SD = .31	concert	0.00	SD = .32	blanket	0.50
	dance	0.58		deep	0.03
	drum	0.70		dream	0.70
	horn	0.03		nap	0.00
	instrument	0.00		peace	0.88
	note	0.30		pillow	0.85
	piano	0.83		relax	0.60
	radio	0.55		rest	0.08
	sing	0.70		snore	0.43
	sound	0.08		tired	0.23
needle	cloth	0.68		yawn	0.78
Mean .49	eye	0.89	sweet	cake	0.92
SD = .35	haystack	0.84	Mean = .54	candy	0.95
	injection	0.24	SD = .39	chocolate	0.34
	knitting	0.58		cookie	0.89
	pin	0.68		sugar	0.00
	point	0.00		heart	0.74
	sewing	0.81		honey	0.42
	sharp	0.00		pie	0.92
	thimble	0.29		soda	0.39
	thorn	0.82		sour	0.00
	thread	0.42		taste	0.03
sheep	bull	0.71		tooth	0.87

Table 3 (continued)

DRM Category and Study List Words		Proportion	DRM Category and Study List Words		Proportion
Mean = .45	country	0.00	thief	bag	0.05
SD = .39	dog	0.68	Mean = .27	cop	0.08
	fence	0.97	SD = .27	gun	0.80
	field	0.45		jail	0.40
	grass	0.00		jewels	0.15
	herd	0.11		law	0.00
	horse	0.92		mask	0.55
	lamb	0.71		money	0.53
	pasture	0.00		movie	0.10
	shear	0.08		steal	0.28
	wool	0.05		villain	0.00

Appendix B

Table 4 DRM words and normed picture labels used in Experiment 1

Critical Word	DRM List Word	Normed Label	Critical Word	DRM List Word	Normed Label
sweet	sour	nuts	chair	table	table
	candy	candy		sit	gymnast
	sugar	ant pile		legs	legs
	taste	tongue		seat	seat
	tooth	tooth		couch	couch
	honey	honey		desk	desk
	soda	can		recliner	recliner
	chocolate	chocolate		cushion	pillow
	heart	hearts		swivel	swivel
	cake	cake		stool	stool
	pie	pie		rocking	seesaw
	cookie	cookie		bench	bench
needle	thread	thread	sleep	bed	bed
	pin	safety pin		rest	lounging
	eye	eye		awake	awake
	sewing	sewing		tired	bored
	sharp	dagger		dream	dream
	point	pencil		blanket	blanket
	thimble	microphone		snore	snoring
	haystack	hay		nap	nap
	thorn	thorns		peace	peace
	injection	shot		yawn	yawn
	cloth	table cloth		pillow	pillow
	knitting	knitting		deep	canyon
fruit	apple	apple	mountain	hill	hill
	vegetable	carrots		valley	farm
	orange	orange		climb	ladder
	kiwi	kiwi		goat	goat
	pear	pear		bike	bicycle
	banana	banana		ski	skiing
	berry	grapes		stream	golf
	cherry	cherries		cave	cave
	basket	basket		cliff	cliff
	salad	food		desert	pyramid
	bowl	bowl		magic	magic
	cocktail	martini		rain	rain

Appendix C: Instructions given in each Experiment

Experiment 1

Study instructions

Auditory only In this part of the experiment you will hear a series of words one at a time.

Please try to remember the words.

Visual word In this part of the experiment you will see a series of words presented one at a time. You will also hear that word. Please try to remember the words.

Picture In this part of the experiment you will see a series of pictures one at a time. You will also hear a related word at the same time. Please try to remember the words and pictures.

Norm word In this part of the experiment you will see a series of words presented one at a time. You will also hear a related word. Please try to remember the words you hear and the words you see.

Recall test instructions In a moment the experimenter will bring you a sheet of paper. Please try to remember as many items as you can from the list you just studied. Please write the items on the sheet of paper. You will have two and a half minutes for recall. Although we want you to remember as much as possible, you should not just guess randomly.

Recognition test instructions In this part of the experiment you will see a list of words one at a time. For each word decide if the word was in the list you studied. Press the "Y" key if the word was in the list. Press the "N" key if the word was not in the list.

Experiment 2

Study instructions

Auditory only During this experiment you will hear a list of words presented one at a time. Please do your best to

try and remember all the words presented. Your memory for this information will be tested later in the experiment

Matched visual word During this experiment you will hear a list of words presented one at a time. You will also see each word on the screen. Please do your best to try and remember all the words presented. Your memory for this information will be tested later in the experiment.

Matched picture and related picture conditions During this experiment you will hear a list of words presented one at a time. You will also see a picture presented along with each word. Please do your best to try and remember all the words and pictures presented.

Recognition test instructions In a moment you will see a list of words shown one at a time. For each word, please decide if the word is one of the words that you just studied. Although we want you to remember as many words as you can, you should not just guess randomly. When each word is presented press the "Y" key for "Yes, this word was in the study list." or the "N" key for "No, this word was not in the list."

Experiments 3 and 4

Study instructions

Word conditions During this experiment you will hear a list of words presented one at a time. You will also see a word on the screen along with each word you hear. Please do your best to try and remember all the words presented. Your memory for this information will be tested later in the experiment.

Picture conditions During this experiment you will hear a list of words presented one at a time. You will also see a picture on the screen along with each word you hear. Please do your best to try and remember all of the words and pictures presented. Your memory for this information will be tested later in the experiment.

Recognition test instructions *These were identical to those in Experiment 2.*

Appendix D

Table 5 Picture-name agreement for stimuli in Experiments 2, 3, and 4. The picture-name agreement information reported here comes from the International Picture Naming Project (IPNP) database (Szekely et al. 2003; Szekely et al., 2004).

Category	Matched Picture		Related Picture	
	Picture Name	Agreement	Picture Name	Agreement
Four-Footed Animal	Cat	.959	Fish	1.000
	Horse	1.000	Saddle	1.000
	Bear	.820	Mountain	.940
	Lion	1.000	Cage	.918
	Tiger	.915	Bone	1.000
	Cow	.938	Bell	1.000
	Elephant	.980	Peanut	1.000
	Deer	.898	Boot	.900
Part of the Human Body	Arm	.837	Violin	.820
	Foot	.980	Hoof	.917
	Finger	.980	Cigarette	.939
	Eye	.980	Glasses	.960
	Hand	.979	Glove	1.000
	Nose	1.000	Clown	1.000
	Ear	1.000	Lightning	.837
	Heart	1.000	Stethoscope	.935
A Fruit	Orange	.959	Sun	1.000
	Banana	1.000	Monkey	1.000
	Grape	.900	Bottle	.898
	Strawberry	1.000	Basket	.980
	Pineapple	.980	Palm Tree	.857
	Watermelon	.980	Ant	.880
	Tomato	.980	Pizza	1.000
	Cherry	.900	Jar	.898
An Article of Clothing	Pants	.896	Zipper	1.000
	Sock	1.000	Drawer	1.000
	Shoe	1.000	Ball	1.000
	Hat	.979	Hook	1.000
	Jacket	.917	Scarf	.980
	Bra	1.000	Girl	.920
	Dress	1.000	Hanger	.898
	Tie	.980	Church	.960
A Type of Vehicle	Bus	1.000	Tire	9.000
	Train	1.000	Whistle	1.000
	Truck	.960	Road	.920
	Motorcycle	.960	Helmet	.960
	Tractor	.870	Potato	.933
	Helicopter	1.000	Cloud	.809
	Tank	.905	Binoculars	1.000
	Sled	.960	Snowman	1.000
A Carpenter's Tools	Saw	1.000	Tree	1.000
	Nail	1.000	Picture	.833
	Screwdriver	1.000	Hinge	.878
	Ruler	1.000	King	1.000
	Wrench	.955	Toilet	1.000
	Screw	.878	Light Bulb	.920
	Ax	.864	Fireman	.940
	Ladder	1.000	Window	1.000
A Kitchen Utensil	Fork	1.00	Spaghetti	.940
	Spoon	1.000	Baby	.940
	Spatula	.857	Hamburger	.840
	Pan	.840	Sink	.958
	Bowl	.980	Football	1.000
	Plate	.940	Table	.980
	Can Opener	.957	Can	1.000
	Cup	.840	Butter	.960

Table 5 (continued)

Category	Matched Picture		Related Picture	
	Picture Name	Agreement	Picture Name	Agreement
A Weapon*	Bomb	.898	Fire	.960
	Sword	.920	Knight	.884
	Rope	1.000	Scissors	1.000
	Cannon	1.000	Pirate	.878
	Chain	1.000	Dragon	1.000
	Whip	.867	Camel	1.000
	Rock	.980	Ring	1.000
	Arrow	.980	Castle	1.000

*For the weapons category, the second highest exemplar, gun, was used as the critical lure item because the highest exemplar, knife, also appeared in the category of kitchen utensils. The substitution was done for the weapons category because this category had more high consistency pictures available

Appendix E

Table 6   Recognition test lure items for Experiments 2, 3, and 4

Critical Lures	Unrelated Lures
Dog	Noun
Pig	Adjective
Sheep	Pronoun
Giraffe	Verb
Leg	Adverb
Toe	Conjunction
Mouth	Preposition
Stomach	Interjection
Apple	Murder
Pear	Rape
Plum	Robbery
Lemon	Theft
Shirt	Assault
Skirt	Stealing
Coat	Arson
Sweater	Burglary
Car	Football
Airplane	Baseball
Bicycle	Basketball
Boat	Tennis
Hammer	Swimming
Level	Soccer
Chisel	Golf
Pliers	Hockey
Knife	Robin
Pot	Sparrow
Stove	Cardinal
Mixer	Blue Jay
Gun	Eagle
Club	Crow
Spear	Bluebird
Fists	Canary

References

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

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms. *Journal of Experimental Psychology, 80,* 1-46.

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to remember than things that were. *Psychological Science, 9,* 484-489.

Cleary, A. M., & Greene, R. L. (2002). Paradoxical effects of presentation modality on false memory. *Memory, 10,* 55-61.

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

Dewhurst, S. A. (2001). Category repetition and false recognition: Effect of instance frequency and category size. *Journal of Memory and Language, 44,* 153-167.

Dodson, C. S., & Hege, A. C. G. (2005). Speeded retrieval abolishes the false memory suppression effect: Evidence for the distinctiveness heuristic. *Psychonomic Bulletin & Review, 12,* 726-731.

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46,* 782–803.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Foley, M. A., Hughes, K., Librot, H., & Paysnick, A. (2009). Imagery encoding effects on memory in the DRM paradigm: A test of competing predictions. *Applied Cognitive Psychology, 23,* 828-848.

Foley, M. A., Wozniak, K., & Gillum, A. (2006). Imagination and false memory inductions:Investigating the role of process, content, and source of imaginations. *Applied Cognitive Psychology, 20,* 1119–1141.

Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology & Aging, 22,* 209–213.

Gallo, D. A., Foster, K. T., & Johnson, E. L. (2009). Elevated false recollection of emotional pictures in younger and older adults. *Psychology and Aging, 24,* 81-88.

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. III. (2001a). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 339-353.

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*, 40, 271-276.

Gallo, D. A., & Roediger, H. L., III. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition, 31,* 1036–1044.

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

Gallo, D. A., Weis, J. A., & Schacter, D. L. (2004). Reducing false recognition with criteria recollection tests: Distinctiveness heuristic vs. criterion shifts. *Journal of Memory and Language, 51,* 473-493.

Ghetti, S., Qin, J. J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology, 38,* 705 – 718.

Gray, S. J., & Gallo, D. A. (2015). Disregarding familiarity during recollection attempts: Content specific recapitulation as a retrieval strategy. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41,* 134-147.

Hege, A. C. G., & Dodson, C. S. (2004). Why distinctive information reduces false memories: Evidence for both reduced relational processing and the distinctiveness heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 787-795.

Hicks, J. L. & Starnes, J. J. (2006). The roles of associative strength and memorability in the contextualization of false memory. *Journal of Memory and Language, 54,* 39-53.

Howe, M. L. (2006). Developmental invariance in distinctiveness effects in memory. *Developmental Psychology, 42,* 1193 – 1205.

Howe, M.L. (2008). Visual distinctiveness and the development of children's false memories. *Child Development, 79,* 65-79.

Howe, M. L., Cicchetti, D., Toth, S. L., & Cerrito, B. M. (2004). True and false memories in maltreated children. *Child Development, 75,* 1402–1417.

Huff, M J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1246-1256.

Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). The effects of distinctive encoding on correct and false memory: A meta-analytic view of costs and benefits and their origins in the DRM paradigm. *Psychonomic Bulletin & Review, 22,* 349-365.

Hunt, R. R., & Einstein, G. O. (1981). Relational 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 & Language*, 32, 421-445.

Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The role of individual item and relational information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 454-464.

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.

Israel, L., & Schacter, D.L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4,* 577-581.

Jacoby, L. L., Shimuzi, Y., Daniels, K. A., & Rhodes, M. J. (2005). Modes of control in recognition and source memory: Depth of retrieval. Psychonomic Bulletin & Review, 12, 852-857.

Johnson, M. K. (2006). Memory and reality. *American Psychologist*, 61, 760–771.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A.(1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, 94**,** 37–64.

Kellogg, R. (2001). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 913-919.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, 37**,** 555–583.

Koutstaal, W., Schacter, D. L., & Brenner, C. (2001). Dual task demands and gist-based false recognition in younger and older adults. *Journal of Memory and Language, 44,* 399-426.

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1997). Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Psychology*, 16, 181-224.

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

McConnell, M. M. & Hunt, R. R. (2007). Can False Memories Be Corrected by Feedback in the DRM Paradigm? *Memory & Cognition, 35,* 999-1006.

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin and Review, 15,* 237-245.

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, 508-520.

Mullet, H. G., & Marsh, E. J. (2016). Correcting false memory: Errors must be noticed and replaced. *Memory & Cognition*, 44, 403-412.

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

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, 35, 261-285.

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

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8,* 385–407.

Schacter, D. L., Cendan, D. L., Dodson, C. S., & Clifford, E. R. (2001). Retrieval conditions and false recognition: Testing the distinctiveness heuristic. *Psychonomic Bulletin & Review, 8,* 827-833.

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.

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.

Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory and Cognition*, 30, 1054-1064.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Smith, R. E., & Engle, R. W. (2011). Study modality and false recall: The influence of resource availability. *Experimental Psychology, 58,* 117-124.

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

Smith, R. E., Hunt, R. R., & Dunlap, K. D. (2015) Why do pictures, but not visual words, reduce older adults' false memories? *Psychology & Aging, 30,* 647-655.

Smith, R.E., Hunt, R.R., Gallagher, M.P. (2008). The effect of study modality on false recognition. *Memory and Cognition, 36,* 1439-1449.

Smith, R. E., Lozito, P., & Bayen, U. J. (2005). Adult age differences in distinctive processing: The modality effect on false recall. *Psychology and Aging, 20,* 486- 492.

Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Category structure and created memories. *Memory & Cognition, 28,* 386–395.

Smith, S. M., Gerkens, D. R., Pierce, B. H., Choi, H. (2002). The roles of associative responses at study and semantically guided recollection at test in false memory: The Kirkpatrick and Deese hypotheses. *Journal of Memory and Language, 47,* 436-447.

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D.,Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods Instruments & Computers*, 35(4), 621-633.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O., Arevalo, A., Vargha, A., Butler, A. C., Buffington, R., & Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, 51(2), 247-250.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29,* 424-432.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289-335.