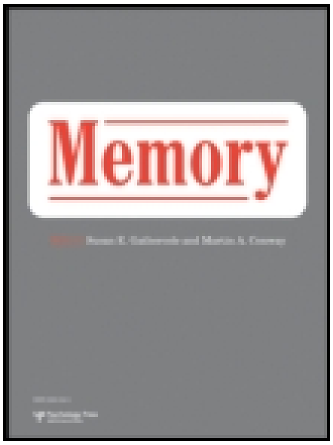

Memory
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pmem20

Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic

David A. Gallo [a] , Deborah M. Bell [a] , Jonathan S. Beier [a] & Daniel L. Schacter [a]
[a] Harvard University , Cambridge, MA, USA
Published online: 17 Feb 2007.

MEMORY, 2006, 14 (6), 730–741

Psychology Press
Taylor & Francis Group

Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic

David A. Gallo, Deborah M. Bell, Jonathan S. Beier, and Daniel L. Schacter
Harvard University, Cambridge, MA, USA

People often use recollection to avoid false memories. At least two types of recollection-based monitoring processes can be identified in the literature. Recall-to-reject is based on the recall of logically inconsistent information (which disqualifies the false event from having occurred), whereas the distinctiveness heuristic is based on the failure to recall to-be-expected information (which is diagnostic of non-occurrence). We attempted to investigate these hypothetical monitoring processes in a single task, as a first step at delineating the functional relationship between them. By design, participants could reject familiar lures by (1) recalling them from a to-be-excluded list (recall-to-reject) or (2) realising the absence of expected picture recollections (the distinctiveness heuristic). Both manipulations reduced false recognition in young adults, suggesting that these two types of monitoring were deployed on the same test. In contrast, older adults had limited success in reducing false recognition with either manipulation, indicating deficits in recollection-based monitoring processes. Depending on how a retrieval task is structured, attempts to use one monitoring process might interfere with another, especially in older adults.

Retrieval monitoring can be defined as the use of retrieved information and expectations to make a memory decision. Successful monitoring leads to the acceptance of true memories and the rejection of false memories that would otherwise be accepted as true. Many types of information could contribute to such decisions, such as source credibility, plausibility assumptions, and knowledge of one's own memory abilities, but a major focus of memory research has been on understanding how recollection itself can inform memory accuracy. How can consciously recalled (or recollected) information help us avoid false memories, and how are these processes affected by stimulus properties, retrieval strategies, healthy ageing, and the like?

It has been proposed that people can use recall or recollection to avoid false recognition through at least two different types of decision process (Gallo, 2004). *Disqualifying* monitoring refers to the rejection of a memory based on the recall of logically inconsistent information (e.g., "I didn't see this item, because I remember hearing it, and items only were presented in one modality"). This type of rejection has been key to many of the exclusion tasks that put recollection and familiarity in opposition (e.g., Jacoby, 1991; Jacoby, Jones, & Dolan, 1998), thereby allowing recollection to override familiarity-based false recognition via a "recall-to-reject" process (e.g., Hintzman & Curran, 1994; Rotello, MacMillan & Van Tassel, 2000). For example, in Jacoby (1991), the recall of a word from one list allowed participants to exclude the word as having been presented in the target list (via a mutual exclusivity rule). Similarly, in Gallo (2004), it was found that

Address correspondence to: David A. Gallo, Psychology Department, University of Chicago, 5848 S. University Ave., Chicago, IL 60637, USA. E-mail: dgallo@uchicago.edu

Preliminary analyses of these data were presented at the 10th Cognitive Ageing Conference (Atlanta, GA) in April 2004. This research was supported by grants NIA AG08441 (DLS) and NIA AG21369 (DAG).

http://www.psypress.com/memory DOI:10.1080/09658210600648506

participants could reject non-studied words when they exhaustively recalled all of the words that were studied from the corresponding category (e.g., ''I couldn't have studied *basil*, because I remember studying *parsley*, *rosemary*, and *thyme*, and only three words were studied per category''). Although the types of recalled information varied across these different tasks, the disqualifying decision process is thought to be the same. Participants rejected the lure because they recalled information that logically precluded it from having been presented in the target context.

Diagnostic monitoring refers to the rejection of a memory based on the failure of recollection to conform to expectations (e.g., ''I didn't see this item, because I would remember it if I had''). This decision process is central to source memory attributions, especially when alternative sources elicit qualitatively different types of recollections, resulting in different expectations as to what should be remembered (e.g., Hicks & Marsh, 2001; Johnson & Raye, 1981; see Ghetti, 2003). For instance, Johnson, Raye, Foley, and Foley (1981, Experiment 2) found that participants were less likely to falsely attribute non-studied words to an internal study source (generating the study words from category cues) than to an external source (simply hearing the words). It was argued that, compared to the external source, the internal source provided more distinctive or elaborate cognitive operations that could later be recollected. The resulting difference in recollective expectations biased false attributions of non-studied events. Because non-studied events were unlikely to elicit recollections of self-generation at study, participants instead attributed them to the external source (the ''it-had-to-be-you'' effect).

Gallo (2004) argued that these sorts of diagnostic monitoring processes might be facilitated by the recall of studied words at test, to the extent that the recall of studied words would make the absence of expected recollections for non-studied words more salient (i.e., a diagnostic recall-to-reject process). As discussed above, exhaustive recall of the studied words did appear to facilitate a disqualifying rejection process in that study, but there was no evidence that recall of only some of the studied words influenced a diagnostic monitoring process. This outcome led to the conclusion that diagnostic monitoring processes instead might be driven by more global expectations formed from the study episode, without having to recall specific studied words at test (e.g., the distinctiveness heuristic, Schacter, Israel, & Racine, 1999). Although it is theoretically possible that specific recalls could influence a diagnostic monitoring process, as could other types of information, these possibilities are not explored here. For the present experiment, we reserve the term ''diagnostic monitoring'' for those instances where global recollective expectations are thought to influence memory decisions, whereas the term ''recall-to-reject'' is used for those decisions in which the recollection of specific information is thought to disqualify a lure from having occurred in a particular context.

In addition to various opposition tasks, source memory tasks, and categorised list tasks, evidence for these two types of rejection processes has been found in the DRM false memory task (Roediger & McDermott, 1995). Participants in this task will often falsely remember a non-studied word (e.g., *sleep*) based on its association with a list of studied words (*bed*, *rest*, *awake* ... etc.). Because this task elicits very high levels of false recognition, it has provided a useful testing ground for studying monitoring processes that can reduce false recognition. Consider first disqualifying monitoring. Gallo, Roediger, and McDermott (2001), and others, have found that forewarning participants about this memory illusion before study can reduce false recognition. These participants were able to figure out the critical lure for some lists during study, and mentally tag them as ''non-studied''. At test, these participants could recall that the critical lures were tagged as non-studied, and thus reject them (see also Neuschatz, Benoit, & Payne, 2003). In this case, the critical lure would be rejected via a recall-to-reject process (i.e., ''This word wasn't studied, because I remember thinking that it was the missing word'').

Disqualifying monitoring was demonstrated in a different way by Dhodia and Metcalfe (1999; see also Smith, Tindell, Pierce, Gilliland & Gerkens, 2001). In their experiments, participants studied two DRM lists, but were later told to reject the items from one of the lists on a recognition test. In a subset of their conditions, they found that false recognition was lower if the critical lure for the target list had been presented in the to-be-excluded list. Dhodia and Metcalfe (1999) proposed that participants had used what could be considered a recall-to-reject strategy to reduce false recognition (i.e., ''This item wasn't studied in the target list, because I remember it from the exclusion list''). This exclusion methodology is very different from the warning methodology, but in both cases participants were thought to have rejected critical lures by recalling, at test, information that was inconsistent with the presentation of these lures in the target list. The recalled information disqualified the questionable event from having occurred in the target context.

Consider next diagnostic monitoring. Schacter and colleagues have argued that participants can reduce DRM false recognition via a process they dubbed the distinctiveness heuristic (Schacter et al., 1999; see Schacter & Wiseman, 2006, for review). For instance, Israel and Schacter (1997) found that participants were less likely to falsely recognise critical lures when each list word was studied with a corresponding picture, compared to a word-only condition. Importantly, these effects were found even when test items were presented only as words (i.e., pictures were not re-presented at test), so that participants had to search memory for picture recollections. To explain these effects Israel and Schacter (1997) argued that participants expected more distinctive perceptual representations after studying pictures, and thus were better able to avoid false recognition of non-studied items (e.g., "This item wasn't studied, because I don't remember seeing a picture for it"). The failure to recall the expected distinctive information was diagnostic of the questionable event's non-occurrence in the context of interest.

Several other lines of evidence favour the distinctiveness heuristic account of these study format effects, as opposed to alternative accounts. For instance, some have argued that studying pictures might impair false recognition by weakening relational processing of the DRM list, thereby reducing the familiarity of the critically related lure (e.g., Arndt & Reder, 2003). At least four sets of findings favour the distinctiveness heuristic account: (1) As predicted by the distinctiveness heuristic, format effects were not found when study format was manipulated within-participants, and thus participants could not expect distinctive recollections for all studied items (Schacter, Cendan, Dodson, & Clifford, 2001; Schacter et al., 1999). A reduced-relational account would still predict a format effect in this situation. (2) As predicted by the distinctiveness heuristic, format effects are often found for false alarms to unrelated lures (Schacter et al., 1999, 2001). A relational account does not apply to unrelated lures. (3) Direct tests of memory for relational information showed that, after studying pictures, relational processing was not sufficiently reduced to account for the obtained reductions in false recognition (Schacter et al., 2001). (4) Studying pictures reduces false recognition in other tasks that do not rely on relational processing to elicit false recognition (e.g., Dodson & Schacter, 2002; Gallo, Weiss, & Schacter, 2004). These last findings are particularly informative. As is the case for a recall-to-reject process, the distinctiveness heuristic appears to be a fundamental monitoring process that generalises across tasks.

THE CURRENT EXPERIMENT

Although there is evidence for both diagnostic and disqualifying monitoring processes across studies, these processes have never been directly compared in a single experiment. With the current experiment we wanted to investigate whether recall-to-reject and the distinctiveness heuristic could be demonstrated on the same test, with the same participants. Finding evidence for these two processes on the same test is important for understanding the potential relationship between the two. Prior studies demonstrate that recollection can enhance discrimination on a variety of recognition tests, but such cross-task comparisons are limited. Not only is the type of monitoring potentially different across tasks, but so too are other retrieval strategies and processes that might influence responding. Thus, although we can draw reasonable theoretical distinctions about monitoring processes, these intuitions need to be tested using a single task that holds all other factors constant. The current experiment was a first step towards that goal.

Of course, even if the two monitoring processes are functionally distinct, it is not a given that they can co-occur. Because these two monitoring processes are thought to rely on qualitatively different types of decision processes (i.e., disqualifying and diagnostic) it is unclear if participants will be able to use both processes to reduce false recognition in the same task. Rejections based on disqualifying monitoring require that participants recollect information that is inconsistent with the occurrence of the lure, and searching for these specific recollections might preclude the use of more general memorability expectations that could facilitate a diagnostic monitoring process. Finding evidence for both processes in the same task would suggest that they are not mutually inconsistent, thereby broadening the extent to which recollection-based monitoring could potentially help to reduce false memories.

The basic DRM procedure was modified to allow us to measure both a recall-to-reject process and the distinctiveness heuristic. In brief, participants heard several DRM study lists, and each word was paired either with the same word visually (Word condition) or with a corresponding picture (Picture condition). This study format manipulation allowed us to investigate the effects of a distinctiveness heuristic on false recognition (Word condition > Picture condition), as in Israel and Schacter (1997) and Schacter et al. (1999). After this study phase, but before the test, participants were warned about related lures. To help them avoid false recognition, they were then presented with a list of some of the related lures in an exclusion list (these lures were presented as words in the Word condition, and as pictures in the Picture condition). They were told to encode (or "tag") these lures as words that were not presented in the study phase, so that they could reject them on the recognition test. This exclusion-list manipulation allowed us to investigate whether participants could use a recall-to-reject strategy to reduce false recognition (untagged lures > tagged lures), as in Dhodia and Metcalfe (1999) and Smith et al. (2001). Following the presentation of the exclusion list, participants took a recognition memory test containing studied words, related lure words (presented in the exclusion list or not), and unrelated control lures (not presented in any prior phase).

If the distinctiveness heuristic and recall-to-reject reflect the operation of different processes, then we would expect different patterns of results according to whether or not lures were presented on the exclusion list. The distinctiveness heuristic is based on the absence of recollection and the resulting diagnostic decision. Thus, in the Picture condition, this heuristic should only reduce false recognition of lures that were not presented in the exclusion list, and hence could not elicit distinctive recollections (i.e., untagged lures and control lures). Tagged lures were presented as pictures in the Picture condition, and so the logic of the distinctiveness heuristic would not apply to these lures. In contrast, a recall-to-reject process depends on recollecting information that is inconsistent with the lure's presentation in the target context and the resulting disqualifying decision. This process should reduce false recognition only for those lures that are recalled from the exclusion list, regardless of the study format (words or pictures). Importantly, study format was held constant within a condition, so that *absolute* differences in perceptual features (e.g., pictures vs. words) could not be used to make the *relative* discrimination between the study phase and the exclusion list in each condition. Because presentation format was held constant, participants instead had to rely on their memory for list membership to discriminate between the study lists and the exclusion list. Assuming that participants used the same list-tagging strategy in each condition, we predicted the same recall-to-reject process in the Word and Picture conditions.

The other goal of this study was to investigate, in a single task, whether ageing differentially affects these two types of recollection-based monitoring. A wealth of research demonstrates that ageing impairs source memory, especially when sources are similar, perhaps attributable to reduced functioning of frontal/medial temporal memory systems (e.g., Glisky, Polster, & Routhieaux, 1995; Henkel, Johnson, & De Leonardis, 1998). Because we equated the perceptual format between the study phase and exclusion list, we predicted that we would find large age-related deficits in the ability to discriminate between these two sources. As a result, older adults should have difficulty using the recall-to-reject strategy, potentially not benefiting at all from presentation in the exclusion list (e.g., Jacoby, 1999).

We did not expect to find an age-related impairment in the ability to use the distinctiveness heuristic. Prior research has shown that older adults are just as likely as younger adults to use the distinctiveness heuristic to suppress false recognition, both in the DRM task (e.g., Schacter et al., 1999) and in other tasks (e.g., Dodson & Schacter, 2002). If ageing impairs recall-to-reject but not the distinctiveness heuristic, then such a dissociation would be strong evidence that these two types of recollection-based monitoring are qualitatively different. Of course, even if the two processes are differentially affected by age, older adults might have difficulty using them both in the same task. Watson, McDermott, and Balota (2004) found that older adults were less likely than younger adults to reduce DRM false recall when two manipulations were combined in a single task (i.e., warnings and practice). If these findings reflect a general depletion of monitoring resources in ageing, then older adults also might have difficulty using the two types of monitoring under scrutiny in our task.

METHOD

Participants

A total of 56 younger adults ($M = 19.4$ years; range 18–24) and 56 older adults ($M = 71.7$ years; range 60–82) participated for course credit or pay. Younger adults were recruited from sign-up sheets in the Harvard Psychology Department, and older adults were recruited from advertisements in the surrounding community. Older adults were screened for depression (Mood Assessment Scale), psychoactive medications, or neurological disease/insults (self-report), and for significant impairment on standard cognitive tests (FAS, digit/symbol, forward & backward span, vocabulary; all obtained in a separate testing session). In general, older adults were cognitively high functioning, and many had participated in other memory experiments (but none involving the DRM task).

Materials and design

A total of 20 DRM lists, modified for pictorial presentation, were drawn from Israel and Schacter's (1997) materials. Each list contained 12 strong associates to a critical lure, with study words arranged in descending associative strength. In addition to the critical lure, a weak lure was identified for each list by drawing an associate from positions 6–12, yielding a total of 11 words in each study list. Ten lists were used for the main study phase (110 words), with the ten non-studied lists providing control lures on the memory test (lists were counterbalanced across studied and non-studied conditions). The exclusion list contained ten lures (hereafter ''tagged lures''), with a critical lure for half the lists and a weak lure for the other half (lists were counterbalanced across critical and weak conditions). The tagging of weak lures ensured that one item corresponding to each of the study lists was presented on the exclusion list, while allowing for half of the critical lures to remain untagged (for comparison to tagged critical lures). Words in the exclusion list were randomly arranged for each participant.

To obtain recognition test items, two list words (from positions 4 and 8) and the two related lures (critical and weak) were drawn from each of the 20 lists, for a total of 80 items. Half of the test words corresponded to the studied lists, including 20 targets (studied words) and 20 related lures (10 presented in the exclusion list, 10 non-presented). The other half of the test words corresponded to the non-studied lists, including 20 target controls and 20 related lure controls (never presented in the exclusion list). Words in the test list were randomly arranged for each participant.

Procedure

Participants were tested in individual rooms. They were told that they would study several lists for an unspecified memory test, and that they should remember what they saw and heard. Auditory labels were presented over speakers that were set to a sufficiently loud volume for each participant, and visual stimuli were presented in the centre of a computer screen. Lists were presented at a rate of 1500 ms per item, with a ''next list'' visual prompt (2500 ms) separating each set of associates. In both conditions, a pre-recorded female voice pronounced the name of each studied word simultaneously with the presentation of the visual stimulus. In the visual word condition, each item was presented in large uppercase letters (black font on a white background). In the picture condition, each item was presented as a black line drawing on a white background. This procedure ensured that both groups received auditory presentation of the word lists, with the only difference being in the distinctiveness of the accompanying visual information (visual words vs. pictures).

After all of the study lists had been presented, participants were given instructions for the test and the exclusion list (referred to as the ''helper list''). They were told that the test would contain studied and non-studied words, and they were warned to avoid false recognition, as follows:

> On the test, some of the ''not studied'' words will seem very familiar because they are related to the studied words, and people often make the mistake of calling these words ''studied''. We want you to try to avoid this mistake. In order to help your memory, I am going to present a helper list before you take the test. This helper list will contain words [and their

pictures] that were not presented in any of the 10 study lists, but that will be on the test. As these words are presented in the helper list, your job is to remember them as words that were not presented during the study phase, so that you won't be tricked into calling them ''studied'' when you take the test. Please pay close attention to what you see and hear, and try to remember that these are all ''not studied'' words. By remembering these helper words as ''not studied'' you will do better on the test.

The experimenter ensured that the participants understood the instructions, and then presented the exclusion list. Within each condition, items in the exclusion list were presented in the same format as items in the main study phase (e.g., auditory labels + visual words in the Word condition, or auditory labels + pictures in the Picture condition). As discussed, this design made source discrimination between the two phases difficult, and it ensured that absolute differences in the perceptual distinctiveness of the stimuli (pictures vs. words) would not differentially benefit source discriminations (and recall-to-reject) in one or the other condition. That is, correctly recalling that an item had been presented in the exclusion list (as opposed to the study phase) should have been equally difficult in the Word and in the Picture conditions.

Following the presentation of the exclusion list, more thorough test instructions were given. Participants were reminded that they were to discriminate between studied and non-studied words from the original study phase (by pressing keys labelled ''S'' and ''NS''). They were also reminded that if they recalled the word from the ''helper'' list (the exclusion list) then they could be sure it was not presented in the original study phase (and should press ''NS''). Finally, they were told that some test words were never presented in either the study phase or the helper list, and that they should press ''NS'' for these words too. Test words were presented visually on the computer screen, and responding was self-paced. Participants were told to take their time to be as accurate as possible.

RESULTS AND DISCUSSION

Results for younger adults are presented first, followed by those for older adults. Unless otherwise discussed, all results reported here were significant at $p < .05$ (two-tailed).

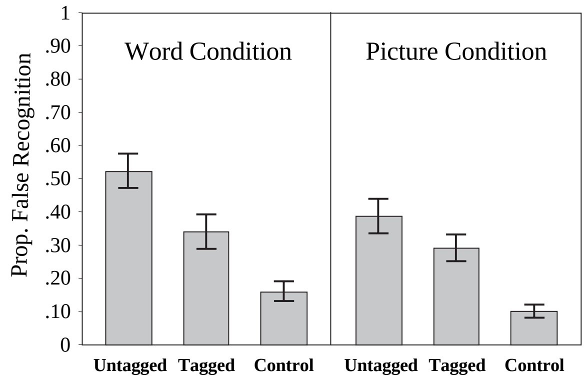

Figure 1. False recognition of critical lures as a function of study format (visual words vs pictures) in younger adults. Untagged lures were related to study lists but never studied, tagged lures were related to the study lists and presented in the exclusion list, and control lures were drawn from non-studied lists (unrelated). Bars represent standard error of the mean.

Younger adults

True recognition of words from the study lists was high (hit rate = .74, collapsing across study formats) and participants could readily discriminate these studied words from their control words (i.e., list words drawn from non-studied lists, false alarm rate = .08), $t(55) = 30.63$, $SEM = .022$. Consistent with prior studies using these materials (e.g., Israel & Schacter, 1997; Schacter et al., 1999), true recognition of list words did not differ across study formats (Word condition mean = .75, Picture condition = .73), $t(54) < 1$, potentially because picture benefits on true memory were offset by the use of words at test.[1]

Figure 1 presents results for critical lures in younger adults. As predicted, tagging related lures in the exclusion list reduced false recognition relative to untagged lures, demonstrating a recall-to-reject strategy. Also as predicted, false recognition was lower in the Picture condition than in the Word condition, demonstrating a distinctiveness heuristic. To confirm these observations, a 3 (item type) × 2 (study condition)

[1] Another possibility is that, because word recollections were less distinctive than picture recollections, participants were more likely to rely on familiarity (or gist-based similarity) at retrieval in the Word condition, thereby boosting hit rates. More generally, it should be noted that picture-superiority effects on hit rates are not always found in between-participants designs (see Mintzer & Snodgrass, 1999).

ANOVA revealed a main effect of item type, $F(2, 108) = 41.21$, $MSE = 0.036$, a main effect for study condition, $F(1, 54) = 4.36$, $MSE = 0.065$, and no interaction, $F < 1$. The main effect of item type indicates that false recognition to untagged lures (mean = .46, collapsed across study conditions) was greater than that to tagged lures (.31), $t(55) = 3.57$, $SEM = .040$, which was greater than that to unrelated control lures (.13), $t(55) = 5.26$, $SEM = .034$. The main effect of study condition indicates that false recognition was lower in the Picture condition (mean = .26, collapsed across all lures) than in the Word condition (.34). Although the expected interaction was not significant, the effect of study condition was numerically larger on untagged lures (13%) than on tagged lures (5%), a point to which we return later.

Figure 2 presents results for the weak lures in younger adults. In contrast to the critical lures, false recognition to weak lures was enhanced by presentation in the exclusion list. A 3×2 ANOVA revealed a main effect of item type, $F(2, 108) = 40.79$, $MSE = 0.021$, no effect of study condition, $F(1, 54) = 1.50$, $p = .23$, and no interaction, $F(2, 108) = 1.12$, $p = .33$. The main effect of item type indicates that false recognition to tagged lures (mean = .31, collapsed across study conditions) was greater than that to untagged lures (.14), $t(55) = 5.43$, $SEM = .031$, which was greater than that to unrelated controls (.07), $t(55) = 2.98$, $SEM = .024$. This reverse tagging effect for weak lures (relative to critical lures) suggests that participants sometimes confused recollections from the exclusion list with the study phase. Because false recognition of untagged weak lures was low, any benefit from presenting weak lures in the exclusion list was outweighed by the added familiarity and source confusion. The low levels of false recognition for weak lures also might have masked the effect of study condition on false recognition, which was in the predicted direction (Word condition > Picture condition) for the non-studied lures (untagged and controls).

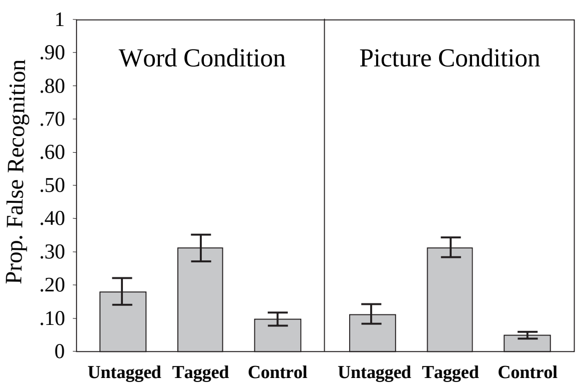

Figure 2. False recognition of weak lures as a function of study format (visual words vs pictures) in younger adults. See Figure 1 caption for details.

We found reduced levels of false recognition to critical lures in the Picture condition (relative to the Word condition), and also to critical lures that were tagged in the exclusion list (in both conditions), but two other predictions were not borne out in the aforementioned analyses. First, the tagged lures should not have benefited from a distinctiveness heuristic. Because tagged lures were actually studied in the exclusion list, participants in the Picture condition should have been able to recollect a picture for these lures. Thus, the logic of the distinctiveness heuristic (i.e., "This item wasn't studied, because I don't recollect a corresponding picture") did not apply to the tagged lures. Inspection of the data in Figures 1 and 2 suggests that the effect of study condition on tagged lures was smaller than that on untagged lures (or non-existent), but the interaction was not statistically significant. Second, we expected that the distinctive heuristic would apply to the weak lures. Although these effects were in the predicted direction, they also were not significant, potentially owing to floor effects. As a more powerful test of these predictions, we pooled the data from critical and weak lures and separately analysed non-studied lures (i.e., untagged lures and control lures) and studied lures (i.e., tagged lures presented in the exclusion list). Based on our a priori hypotheses, the distinctiveness heuristic should have affected all non-studied lures, but should not have affected tagged lures.

A 2 (item type) $\times$ 2 (lure strength) $\times$ 2 (study condition) ANOVA on all non-studied lures revealed a main effect of item type, $F(1, 54) = 96.18$, $MSE = 0.023$, demonstrating that related lures were more likely to be falsely recognised than unrelated controls, and a main effect of lure strength, $F(1, 54) = 71.61$, $MSE = 0.028$, demonstrating that strong lures were more likely to be falsely recognised than weak lures. There also was an interaction between these two variables, $F(1, 54) = 40.36$, $MSE = 0.022$, confirming that the relatedness effect on false recognition was greater for critical lures than for weak lures. Most important, there was a main effect of study condition, $F(1, 54) = 5.94$, $MSE = 0.057$, and this variable did not interact with any others (all ps > .25). As predicted, false alarms to all unrelated lures were reduced in the more distinctive study condition (Word condition > Picture condition).

In contrast to the non-studied lures, a 2 (lure strength) × 2 (study condition) ANOVA on all tagged lures resulted in no effect of strength, $F(1, 54) < 1$, no effect of study condition, $F(1, 54) < 1$, and no interaction, $F(1, 54) = 1.09$, $p = .30$. As predicted, participants were just as likely to falsely recognise tagged lures in the Word condition (mean = .33) as in the Picture condition (.30), indicating that these lures did not benefit from a distinctiveness heuristic. By equating the presentation format between the study lists and the exclusion list, we were effective at equating the discrimination between these two sources across the Word and Picture conditions.

Older adults

As was the case with younger adults, true recognition of studied words was much higher than false recognition to control words in older adults (means = .72 and .04, collapsing across formats), $t(55) = 29.91$, $SEM = .023$, demonstrating very good discrimination. Unlike younger adults, true recognition was greater in the Word condition (.78) than in the Picture condition (.66) for older adults, $t(54) = 2.65$, $SEM = .042$. There was no a priori reason to expect such an effect, and to ensure that this difference did not affect our other findings, we excluded six participants from each condition to equate true recognition across the Word and Picture conditions (both hit rates = .72, both control FAs = .04). The patterns of all other results and conclusions were identical for this matched group as for the entire group, so we report results from the entire dataset below.

Figure 3 presents results for critical lures in older adults. Relative to younger adults, older adults were impaired in their use of both a recall-to-reject strategy and a distinctiveness heuristic. A 3 (item type) × 2 (study condition) ANOVA resulted in a main effect of item type, $F(2, 108) = 102.36$, $MSE = 0.032$, but no effect of study condition and no interaction, both Fs < 1. The effect of item type indicates that false recognition of related lures (tagged and untagged) was greater than false alarms to control lures (both ps < .001), but false recognition did not differ for tagged lures (mean = .51, collapsed across conditions) and untagged lures (.47), $t(55) = 1.30$, $SEM = .030$, indicating a deficit in recall-to-reject.

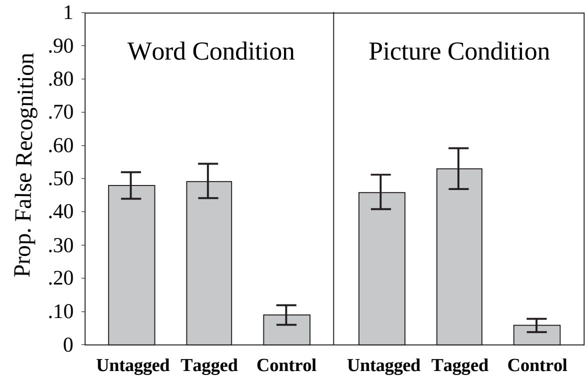

Figure 3. False recognition of critical lures as a function of study format (visual words vs pictures) in older adults. See Figure 1 caption for details.

This is not to say that older adults were not trying to exclude tagged items. If they were not trying, and had simply treated tagged lures like other studied items, then false recognition of tagged lures (.51) should have been greater than false recognition of untagged (non-studied) lures (.47), and as high as true recognition of other studied items (.72). Instead, these data suggest that older adults had tried to exclude these items, but were less successful than were younger adults.

Figure 4 presents results for the weak lures in older adults. A 3 (item type) × 2 (study condition) ANOVA again revealed a main effect of item type, $F(2, 108) = 61.51$, $MSE = 0.037$, but no effect of study condition, $F(1, 54) = 1.44$, $p = .24$, and no interaction $F(2, 108) < 1$. False recognition for tagged weak lures (mean = .44, collapsed across study condition) was greater than that to untagged weak lures (.19), $t(55) = 5.85$, $SEM = .043$, which was greater than that to unrelated control lures (.04), $t(55) = 6.11$, $SEM = .025$. As was the case with younger adults, these effects

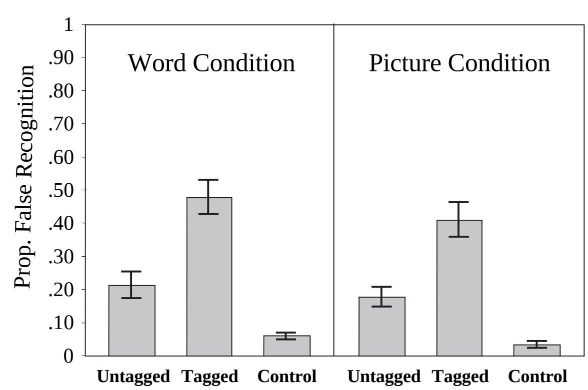

Figure 4. False recognition of weak lures as a function of study format (visual words vs pictures) in older adults. See Figure 1 caption for details.

demonstrate that presenting lures in the exclusion list enhanced familiarity or source confusions, leading to an overall increase in false recognition of weak lures.

We also conducted separate ANOVAs on non-studied lures and studied lures in older adults. A 2 (item type) × 2 (lure strength) × 2 (study condition) ANOVA on all non-studied lures revealed a main effect of item type, $F(1, 54) = 159.63$, $MSE = 0.026$, a main effect of lure strength, $F(1, 54) = 69.64$, $MSE = 0.019$, and an interaction between these two, $F(1, 54) = 35.98$, $MSE = 0.023$. As in the younger adults, these effects indicate that false recognition was greater for related than for unrelated lures, and that this relatedness effect was greater for critical lures than for weak lures. More important, there was no effect of study format, $F(1, 54) < 1$, and this variable did not interact with any of the others, all Fs < 1. Unlike the younger adults, there was no evidence that a distinctiveness heuristic had influenced non-studied lures in older adults. A 2 (lure strength) × 2 (study condition) ANOVA on all tagged lures revealed no significant effects and no interaction, all ps $> .16$, demonstrating that false recognition of tagged lures was equivalent in the Word condition (mean = .48, collapsed across strength) and in the Picture condition (.47).

GENERAL DISCUSSION

We found that younger adults could use both recall-to-reject and the distinctiveness heuristic to suppress false recognition on the same test. To our knowledge, this is the first demonstration that these two types of monitoring can be used on the same test, allowing us to directly compare the underlying processes. In younger adults, the data were consistent with the idea that the two recollection-based monitoring processes depended on qualitatively different types of decisions. The distinctiveness heuristic is a diagnostic monitoring process, based on the absence of recollection, and so it should apply only to lures that were never studied in any phase (and hence could not elicit recollection). Consistent with this prediction, false recognition of untagged lures and non-studied control lures differed across the format conditions (Word > Picture), but false recognition to lures that were studied in the exclusion list did not differ (or differed less). In contrast, the recall-to-reject strategy should have selectively affected tagged lures, because these were the only lures for which disqualifying information (presentation in the exclusion list) could be recalled. Consistent with this prediction, we found evidence for recall-to-reject (untagged lures > tagged lures) for critical lures in both study conditions. If anything, the tagging effect was larger in the Word condition (18%) than in the Picture condition (10%), which is contrary to what one would expect if recollective distinctiveness had benefited a recall-to-reject strategy more than the distinctiveness heuristic.

The finding that older adults were less able than younger adults to use the exclusion list to suppress false recognition was not surprising, given the expansive literature on source memory and recall-to-reject deficits in older adults. However, our study is the first to report age-related deficits in the use of the distinctiveness heuristic. One possible explanation is that our older adults did not encode distinctive visual information from pictures at study, as reflected in the reduced hit rate to studied items in the Picture condition. We cannot definitively rule out this possibility, but note that there was no evidence for a distinctiveness heuristic in older adults even when we compared a subset of participants that were equated on hit rates across study conditions. Further, we used the same study materials and general presentation procedures as in Schacter et al. (1999), who showed a robust distinctiveness heuristic in a group of older adults that was drawn from the same participant pool as in the current study (see also Budson, Sitarski, Daffner, & Schacter, 2002). These previous findings suggest that there was something different about our task that interfered with the use of a distinctiveness heuristic at retrieval, as opposed to impoverished encoding of pictures at study.

The key procedural difference between our procedure and these other studies was that we warned participants about the false recognition effect (before test) and provided them with a list of to-be-excluded lures. Attempts to use the exclusion list to reduce false recognition, via a recall-to-reject strategy, might have interfered with older adults' ability to use the distinctiveness heuristic. The ability to reject lures from the exclusion list was lower in older adults (mean rejection = .52, collapsing across all tagged lures) than in younger adults (.69), $t(110) = 4.21$, $SEM = .039$, and many of our older adults noted in post-experiment comments that the "helper" list (the exclusion list) was more confusing than helpful. Because this source discrimination was so difficult, older adults may have focused all of their strategic efforts on determining whether the lures had been presented in the exclusion list. If a lure was highly familiar but they could not remember its source, then they might have arbitrarily attributed it to the study phase without attempting to use a distinctiveness heuristic. Put differently, older adults might have been so focused in their attempts to apply the disqualifying strategy that they failed to use distinctive recollective expectations to inform diagnostic monitoring processes. This conclusion implies that some monitoring process might actually enhance false memories, by blocking more useful retrieval editing strategies.

This interference interpretation is consistent with the idea that some recollection-based exclusion strategies can be relatively slow and resource demanding (see Yonelinas, 2002), but it is only speculative and awaits further testing. Regardless of the explanation, our results indicate that ageing can cause monitoring impairments that extend beyond recollection-based exclusion (or recall-to-reject). The distinctiveness heuristic is different from recall-to-reject, in that it is based on the failure to recollect expected information (e.g., a picture). Although older adults are just as likely as younger adults to use this type of monitoring in some situations (as demonstrated in prior studies), our results indicate that older adults can be impaired in the use of a distinctiveness heuristic. In this sense our findings are like those from the false fame task, in which it has been demonstrated that older adults are less likely than younger adults to use a source-based exclusion strategy when the test does not make this strategy salient, compared to when it does (e.g., Multhaup, 1995). Older adults may have failed to use the distinctiveness heuristic in our task because the test instructions focused them on other components of the task.

This last observation raises an important question about the generality of our results. If our task focused older adults too much on a disqualifying recall-to-reject strategy, then it might be possible to create a task that emphasises diagnostic monitoring processes instead. By explicitly focusing participants on the distinctiveness of their recollections at test, older adults might show intact use of the distinctiveness heuristic even if a recall-to-reject strategy could also be used. We currently are exploring this idea using the criterial-recollection task (see Gallo et al., 2004), which forces participants to directly query their memory for the to-be-recollected information (e.g., words or pictures). Exploring the interaction of these two monitoring processes in other tasks would also avoid some of the interpretative ambiguities that are inherent in the DRM task, such as whether monitoring occurs during study or test, or whether distinctive study conditions interfere with associative or gist-based encoding (e.g., Arndt & Reder, 2003; Hege & Dodson, 2004).

Frameworks versus processes

In the Introduction we mentioned two research areas that are relevant to recollection-based monitoring processes. Previous demonstrations of disqualifying monitoring processes, such as exclusion-based recall-to-reject, have often been interpreted within a dual-process framework of memory. This framework proposes that recollection and familiarity are qualitatively different types of retrieval, and that participants can strategically use recollection to control the influences of familiarity (e.g., Yonelinas, 2002). Demonstrations of diagnostic monitoring processes, like the distinctiveness heuristic, have often been interpreted within the source-monitoring framework (e.g., Johnson, Hashtroudi, & Lindsay, 1993). This framework emphasises that all memory decisions involve the attribution of retrieved information to a source, based on a variety of decision processes, including those that take retrieval expectations into account. Other frameworks offer other ways to conceptualise recollection-based rejections, such as fuzzy trace theory (e.g., Brainerd, Reyna, Wright, & Mojardin, 2003; Lampinen, Odegard, & Neuschatz, 2004). This framework is akin to the dual-process framework, with the exception that fuzzy trace theory makes specific assumptions about the information content of the memory traces (i.e., gist vs verbatim) that give rise to different types of subjectivity (i.e., familiarity, recollection, and phantom recollection).

Rather than advocating any particular framework or model of memory, our goal here and elsewhere has been to try to characterise the basic processes of recollection-based monitoring. This approach has the benefit of explaining a variety of empirical findings as originating from only a small set of candidate processes, without making many assumptions about the underlying structure of the memory system(s) that gives rise to these processes. One challenge to this approach, though, is that some phenomena might not be readily described by the processes under consideration. As an example, consider the idea of ''recollection-rejection'' that has been proposed as a way of conceptualising recollection-based monitoring processes in fuzzy trace theory (Brainerd et al., 2003). Under this view, the retrieval of a verbatim trace (e.g., *snow*) can facilitate the rejection of a gist-consistent lure (e.g., *sleet*) via the subjective experience of a non-identity judgment (e.g., ''*Sleet* is probably familiar only because I studied the related word *snow*''). In this situation, it is unclear whether the participant is making a disqualifying decision (e.g., ''*Sleet* couldn't have been studied, because I recall *snow*, and I only studied one associate''), a diagnostic decision (''I can vividly recollect *snow*, but not *sleet*, so *sleet* probably wasn't studied''), or some combination of both. Whether a monitoring process is best characterised as disqualifying or diagnostic will depend not only on the specifics of the information that is retrieved, but also on the assumptions that the rememberer makes about the situation and how these assumptions influence the memory decision.

REFERENCES

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

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False memory editing in children and adults. *Psychological Review*, *110*, 762–784.

Budson, A. E., Sitarski, J., Daffner, K. R., & Schacter, D. L. (2002). False recognition of pictures versus words in Alzheimer's disease: The distinctiveness heuristic. *Neuropsychology*, *16*, 163–173.

Dodhia, R. M., & Metcalfe, J. (1999). False memories and source monitoring. *Cognitive Psychology*, *16*, 489–508.

Dodson, C. S., & Schacter, D. L. (2002). Ageing and strategic retrieval processes: Reducing false memories with a distinctiveness heuristic. *Psychology & Ageing*, *17*, 405–415.

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *30*, 120–128.

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

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

Ghetti, S. (2003). Memory for non-occurrences: The role of metacognition. *Journal of Memory and Language*, *48*, 722–739.

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

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

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

Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old–new recognition. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *27*, 375–383.

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

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

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

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

Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review*, *5*, 705–709.

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

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.

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.

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *30*, 332–342.

Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support of the distinctiveness model. *American Journal of Psychology*, *112*, 113–146.

Multhaup, K. S. (1995). Ageing, source, and decision criteria: When false fame errors do and do not occur. *Psychology and Ageing*, *10*, 492–497.

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, Cognition*, *29*, 35–40.

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

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language*, *43*, 67–88.

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.

Schacter, D. L., & Wiseman, A. L. (2006). Reducing memory errors: The distinctiveness heuristic. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory,* (pp. 89–107). New York: Oxford University Press.

Smith, S. M., Tindell, D. R., Pierce, B. P., Gilliland, T. R., & Gerkens, D. R. (2001). Source memory failure in episodic confusion errors. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *27*, 362–374.

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: Assessing the combined influences of practice and warnings in young and old adults. *Memory & Cognition*, *32*, 135–141.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.