Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

False memory-guided eye movements: insights from a DRM-Saccade paradigm

Lauren Knott, Damien Litchfield, Tim Donovan & John E. Marsh

MEMORY
https://doi.org/10.1080/09658211.2024.2307921

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

False memory-guided eye movements: insights from a DRM-Saccade paradigm

Lauren Knott[a], Damien Litchfield[b], Tim Donovan[c] and John E. Marsh[d,e]

[a]Department of Psychology, City, University of London, London, UK; [b]Department of Psychology, Edge Hill University, Ormskirk, UK; [c]Institute of Health, University of Cumbria, Lancaster, UK; [d]School of Psychology and Humanities, University of Central Lancashire, Preston, UK; [e]Department of Business Administration, Technology and Social Sciences, Luleå University of Technology, Luleå, Sweden

ABSTRACT
The Deese-Roediger and McDermott (DRM) paradigm and visually guided saccade tasks are both prominent research tools in their own right. This study introduces a novel DRM-Saccade paradigm, merging both methodologies. We used rule-based saccadic eye movements whereby participants were presented with items at test and were asked to make a saccade to the left or right of the item to denote a recognition or non-recognition decision. We measured old/new recognition decisions and saccadic latencies. Experiment 1 used a pro/anti saccade task to a single target. We found slower saccadic latencies for correct rejection of critical lures, but no latency difference between correct recognition of studied items and false recognition of critical lures. Experiment 2 used a two-target saccade task and also measured corrective saccades. Findings corroborated those from Experiment 1. Participants adjusted their initial decisions to increase accurate recognition of studied items and rejection of unrelated lures but there were no such corrections for critical lures. We argue that rapid saccades indicate cognitive processing driven by familiarity thresholds. These occur before slower source-monitoring is able to process any conflict. The DRM-Saccade task could effectively track real-time cognitive resource use during recognition decisions.

ARTICLE HISTORY
Received 12 October 2023
Accepted 8 January 2024

KEYWORDS
DRM paradigm; false memory; eye saccades; Recognition accuracy

Memory's susceptibility to errors has been well-established through extensive research. A prominent tool for studying these errors is the Deese/Roediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995). This laboratory-based method involves participants studying lists of words, such as "tired", "awake", "quiet", and "bed", which are all related to a non-presented critical lure, for instance, "sleep". A "false memory" is noted when participants recall or recognise this critical lure despite its absence in the list. Remarkably, the frequency of falsely recalling or recognising this lure can mirror that of the genuinely studied words (Roediger & McDermott, 1995). One prevailing explanation for this phenomenon, termed activation/monitoring theory (Roediger et al., 2001), posits that when an item is studied, it can activate related but non-presented items in the mental lexicon due to the spreading activation of conceptual representations. The strength of activation of the related but non-presented items increases the difficulty of making diagnostic source-monitoring decisions (Roediger & McDermott, 2000) about the presence or absence of that item in the list which leads to false recognition or recall.

The robustness of false memory production using the DRM paradigm and its resistance to elimination has been extensively documented since 1995. One focus of this research has been on the seemingly indistinguishability of the false memory for the critical lure and the true memory for the studied list item. For example, remember/know judgments (Roediger & McDermott, 1995) have revealed similarities between list and critical lure items suggesting similar experiences of recognition, detailed features of studied items have become attached to the false memory of the related lure (e.g., content borrowing; see Lampinen et al., 2008; Lyle & Johnson, 2006) and participants have expressed distinct recollections of having heard the critical lure word, knowing which speaker (when there was more than one) presented the critical lure item or what position the word appeared (e.g., Lampinen et al., 1999; Neuschatz et al., 2001; Norman & Schacter, 1997). By contrast, neurocognitive studies have shown that false memories can be differentiated from veridical memories. For example, in examining event-related potentials, Miller et al. (2001) found that false recognition of critical lures produced substantially shorter P300 latencies than

CONTACT Lauren Knott lauren.knott.1@city.ac.uk Department of Psychology, City, University of London, London EC1V 0HB, UK; Damien Litchfield litchfd@edgehill.ac.uk Department of Psychology, Edge Hill University, Ormskirk, Lancashire L39 4QP UK

correct recognition of studied words. Schacter and Slotnick (2004) found that memory critical lures showed greater activity in the ventromedial prefrontal cortex, associated with post-retrieval monitoring, and comparatively reduced activity in the parietal cortex and the parahippocampal gyrus of the brain, associated with visual and auditory sensory contextual reactivations. Whilst the aforementioned studies pertain to differentiation in true and false memories, we note that this is not always the case for neuroimaging studies (see McDermott et al., 2017; Schacter et al., 2012).

Coane et al. (2007) contended that true and false memories can indeed be differentiated, though not all metrics may capture these distinctions effectively. Even if participants label a critical lure as "old", equating its accuracy with a studied list item, discrepancies between these items might still arise. Supporting neurocognitive distinctions above, research has shown that participants response latencies are consistently slower when making a false-positive recognition of a critical lure compared to a correct-hit recognition of a studied list item (Coane et al., 2007; Jou et al., 2004). This evidence suggests that response latencies might better reflect the disparities in processing list and critical lure items. Jou et al. (2004) argued that the explicit response of making an old/new decision in a recognition test may not discriminate between a true and false memory at the conscious level, but the involuntary behavioural manifestation of a slower response to critical lure indicated a discrimination of the false memory from the true memory at an unconscious level. Coane et al. (2007) explored how the activation/monitoring theory (Roediger et al., 2001) might manifest in recognition latencies. Words with strong activation from the encoding phase, such as studied words and critical lures, should be recognised more quickly. However, a lengthier monitoring procedure would further assess this recognition. Words that surpass a high familiarity threshold would prompt swift recognition (e.g., typical for studied words). In contrast, words below a low familiarity threshold would prompt a quick non-recognition response (e.g., as seen with unrelated lures). Words that fall in between these thresholds would undergo source-monitoring, resulting in extended latencies. Aligning with these assumptions, Coane et al. (2007) discovered that correctly rejecting critical lures took longer than rejecting weakly associated lures, suggesting a rapid low threshold decision for weakly related items but not for critical lures. Similar to Jou et al. (2004), falsely recognising critical lures was also slower than correctly recognising studied list items (see also Brown et al., 2000; Jou et al., 2017 for further consistent findings). However, not all studies that have examined response latencies using the DRM paradigm have shown this clear distinction. A study by Thomas and Sommers (2005) examined age-related differences in response latencies for false memories. They reported no statistical difference between list item hits and critical lure false alarms in both their age groups (although Tun et al., 1998, only found this effect for older adults). Hancock et al. (2003) found faster (or similar) response latencies to critical lure false alarms compared to list item hits using a lexical decision task. Here, researchers have argued that the false memory latency demonstrated a strong degree of activation, especially when relying on theme-based strategies to support recognition of list and critical lure items. Note that there are several methodological differences that may account for the differing outcomes reported here, but the explanation provided by Hancock et al. does seem to fit with theoretical arguments posed by Jou et al. (2017). They argued that as the list length increased, the strength of false alarm increased, marked by increasing response rate and decreasing reaction time and was a result of a switch of reliance from a verbatim dominant memory mode to a gist dominant memory mode. The above research suggests that strength of activation is important for recognition latency and there may be additional monitoring processes that impact false memory latencies.

Although Jou et al. (2017) argued that recognition latencies highlight more subtle differences in real and false memories, not reflected in a yes/no or correct/ incorrect responses, a response time to make this binary decision still necessitates an explicit reaction, where any certainty or uncertainty about recognition is solely represented by the duration it takes to respond via a keyboard. Being able to capture ongoing cognitive processes prior to the output/explicit response could highlight further discrepancies/similarities between real and false memories, particularly regarding the role of familiarity and the more controlled recollection/source discrimination. The two experiments presented in this study aim to assess both correct and false memory decisions through eye-movement saccades.

Our methodology draws inspiration from the renowned pro and antisaccade tasks (see Hutton, 2008; Munoz & Everling, 2004 for an overview). In simple terms, saccades represent decisions: decisions about where to direct the eye and whether the movement is essential (Hutton, 2008). Rayner (2009) posits that real-time cognitive activities can be gauged using eye-movement data however, according to Seideman et al. (2018), few studies have directly linked saccade metrics themselves to underlying decision-related processes. In saccade tasks, participants are shown a central fixation cross and after a brief period of time a target dot appears in the periphery either to the left or to the right of the central cross. Participants are asked to look at this target dot as quickly as possible whilst their eye movements are recorded. This is a prosaccade. They are considered to be fast and reflexive, producing very little error. In the antisaccade task participants are presented with the identical stimuli but are told to avoid the prepotent response of looking towards the target dot, and instead to look in the opposite direction (for instance, if the target appears to left, participants must move their eyes to the right, where there is no visible stimulus). They require more cognitive effort. The success of the performance is validated through eye-tracking data. This data specifically captures the direction, speed, and accuracy of eye movements. Successful execution is indicated by the participant's ability to quickly and accurately divert their gaze in the correct direction.

In the eye movement literature, there are a number of memory-guided saccade tasks that examine our ability to fixate a location under different memory loads (Rayner, 2009). For example, the mixed-saccade task is a rule-based version of the antisaccade task, whereby, if the central fixation dot is blue then participants are asked to make a prosaccade response; if red then they are asked to make an antisaccade response (Cherkasova et al., 2002; Rivaud-Pechoux et al., 2007). Both the prosaccade and antisaccade eye movements evolve into controlled, memory-influenced decisions that adhere to the rule specified for that particular trial. The DRM-Saccade paradigm we outline here provides a novel manipulation of rule-based saccade tasks in which participants must make a specific eye movement if they do or do not recognise a centrally presented word on the screen. Rather than employing a colour target to dictate the required prosaccade or antisaccade, we lean on participants' recall of the previously encountered DRM list items (and associated critical lures) to determine their eye movement response.

Models of saccade generation provide further insights into factors (and potential underlying cognitive influences) that contribute to the decision to make a saccade. Race models (e.g., Hanes & Carpenter, 1999; Hanes & Schall, 1995; Logan et al., 1984; Osman et al., 1986) offer a framework for understanding these eye movements. They suggest that saccades are initiated when accumulated evidence for a movement decision reaches a certain threshold, involving a competition between reflexive responses (prosaccade) and cognitive control (antisaccade). Computational models support this by showing how activation from external stimuli and internal control signals accumulate before triggering a saccade (Cutsuridis et al., 2007; Trappenberg et al., 2001; Wilimzig et al., 2006). If cognitive control prevails, an intentional saccade is made; if not, a reflexive, possibly erroneous movement occurs. Additionally, feedback mechanisms continuously monitor and, if necessary, initiate corrective saccades to align the gaze with the intended target, although such corrections are not instantaneous and involve a latency period (Crawford et al., 2005).

The saccadic task provides an opportunity to study the ongoing cognitive processes when making recognition decisions. If we consider models of saccade generation, we can consider how accumulated evidence creates a recognition decision that may be accurate or inaccurate. The prevalent dual-process view holds that implicit spreading activation, which is the source of false memory, is automatic and fast (Jacoby, 1991; Underwood, 1965), but monitoring, which is the basis of a correct rejection, is an effortful, slow, controlled process (Jacoby, 1991; Roediger et al., 2001). Johnson et al. (1994) argued that whilst old–new recognitions can be based on the amount of associative spread of activation, source-monitoring requires more differentiated information. That is, decision mechanisms can use relatively undifferentiated information for old–new recognition decisions but require more differentiated information for source-monitoring decisions. Sufficient information for old–new discrimination becomes available before information for source-monitoring. Therefore, how fast saccade latencies are towards or away from the target will indicate low or high thresholds of familiarity of the item. The speed of antisaccades will indicate the additional time needed to reject the familiar distractor after additional differentiating information supports such decision. Whether or not the correct rejection takes place will depend on whether the familiarity of the lure provides enough evidence to prevent the onset of additional monitoring which would lead to a longer antisaccade or a correction.

To summarise, while numerous studies have delved into the encoding factors influencing false memory, the retrieval aspect remains less explored. Moreover, there is surprisingly little research on how methods of retrieval may help us understand the decision-making criteria employed. Conventionally, participants explicitly signal their recognition or non-recognition of test words through verbal, written, or button-press responses. The present set of experiments required participants to make recognition/non-recognition decisions by making saccadic eye movements towards or away from peripheral targets. Eye-movements are faster and more implicit, capturing ongoing decision-making (Rayner, 2009). Our primary objectives were twofold: first, to discern whether this response modality yields findings consistent with earlier response accuracy and latency studies (e.g., Coane et al., 2007; Jou et al., 2004; Tun et al., 1998), and second, to determine if the DRM-Saccade paradigm reveals differences in corrective saccadic behaviour. This is particularly the case when initial rapid recognition decisions are later modified – either improved or worsened – based on subsequent cognitive processing.

Experiment 1

In our first experiment, we presented a single peripheral target. Participants were guided to look towards this target if they recognised the word and to look away, executing an antisaccade, if they did not. With only one peripheral target, a natural inclination arises to saccade towards it. The central query here is: can this inherent response bias distinguish between true and false memory recognition? To signify recognition of a previously studied item, participants are prompted to direct a prosaccade towards the target, and for non-recognition, an antisaccade away from it.

Experiment 1 first establishes how people respond to the DRM-Saccade task, in terms of recognition rate and saccadic latencies for correct and incorrect decisions. We assessed recognition saccades related to list items, critical lures, and unrelated lures. We analysed saccadic latencies associated with hits for studied items alongside correct rejections and false alarms concerning critical lures and unrelated lures. By evaluating saccadic latencies for correct rejections, we gauged the cognitive effort required to accurately dismiss a critical lure or an unrelated filler lure. Furthermore, a comparison of saccadic latencies for hits against those for critical lure false alarms provided insights into the cognitive processes underpinning true versus false recognition decisions. For this experiment, participants were instructed to look towards the target when they recognised the word (execute a prosaccade) and to look away when they did not (perform an antisaccade). Our primary goal in employing the DRM-Saccade task is to investigate whether it can effectively distinguish between accurate and false recognition decisions, both in terms of response accuracy and latency. Although our approach is exploratory in nature, we hypothesise that items triggering a lower threshold of familiarity will elicit quicker saccadic responses. This is based on the premise that such items demand less cognitive effort for decision-making. In contrast, we anticipate that critical lures, which present a higher familiarity threshold, will require more time for cognitive processing. This is due to the need for more intricate evaluation to understand why these items seem familiar yet are not true memories. The key question is whether the saccadic latencies for correctly recognising list items and falsely recognising critical lures will show a significant difference. This outcome may hinge on the familiarity threshold: if it's sufficiently high, it might prompt a more automatic, reflexive response, whereas a lower threshold could necessitate more deliberate, effortful decision-making.

Method

Participants

Forty participants (32 female) 18–27 years old ($M = 19.00$, $SD = 1.89$) from Edge Hill University, England were recruited for Experiment 1. An a priori power analysis (using G*Power 3.1 software; Faul et al., 2007) was conducted for a suitable sample for a repeated measures design with 3 conditions pertaining to the item type. We used a medium effect size $f = .25$, alpha level of 0.05, power of 0.80, and a conservative correlation estimate of 0.50. This indicated a sample size of 28. As this was a proof of concept study and we wanted to ensure sufficient sample size in case any data had to be removed, we used a sample of 40. All participants had normal or corrected-to-normal vision and spoke English as their first language. They were rewarded with either university course credit or monetary compensation. Ethical approval was received for both Experiments by Edge Hill University Ethics Psychology Sub-Committee.

Design

The memory-guided saccade task required saccade recognition decisions to items from the DRM lists (studied words), the non-presented critical lure, and unrelated filler items. We examined the following dependent measures for these items.

Recognition rate

This is whether participants made the appropriate memory-guided saccade towards or away from the target dot when presented with either the studied word, unrelated filler item, or critical lure. This metric is the most comparable to recognition rates in standard DRM recognition tasks.

Saccadic latency (ms)

This is the time taken for the first eye movement towards the peripheral target (or in the opposite direction to the target) since the test word was presented. As the target dot was presented 8° to the left/right of the central test word, an eye movement was considered a decision (recognised or not recognised) if it was greater than 2°. In addition, consistent with the antisaccade literature (Fischer, 1987), only latencies greater than 80 ms were included, as latencies quicker than 80 ms are considered anticipatory.

Stimuli and apparatus

Twenty-four DRM lists were utilised in Experiment 1 (see Appendix). Lists were either taken from Roediger et al. (2001) or were constructed following the same procedure used by Roediger et al., with associative strength indexes taken from the South Florida Free Association Norms (Nelson et al., 1999). All lists contained 12 items and were matched for Backwards Associative Strength ($M = 0.24$). An additional thirty-two unrelated lures were selected from unused associative lists from the above two sources, ensuring no associative relations to the studied words or critical lures. Therefore, all nonpresented distractor words except the critical lure were unrelated distractors (recognition test filler items). All words were presented on a 19-in. CRT monitor (1024 × 768 pixels, 120 Hz). Eye movements were recorded using an EyeLink 1000 desktop eye tracker (SR Research Ltd, Mississauga, Canada) and stimuli were presented via Experimental Builder software. Participants were seated 57 cm away from the computer screen and used a chin support to maintain this distance. A 9-point calibration was used and calibration was only accepted if the average calibration error was less than 0.5°.

Procedure

An information sheet and consent form were reviewed and signed before commencing with the study. Study lists were presented in blocks of 4 and afterwards participant's memory was tested based on the 4 lists they had just seen. This resulted in 6 blocks in which participants undertook the memory-guided saccade task. This method was utilised to reduce fatigue in the saccade task and all blocks were fully counterbalanced to ensure that lists were seen equally often in each grouped block position. Although we utilised a study-test block method we did not expect to see any decline in false memories across these six blocks. Previous research has shown that without any feedback on the purpose of the test, false recognition for DRM critical lures does not decline (see Jou & Foreman, 2007, with 16 study-test trials and no decline).

Study phase

For each DRM list, participants were presented with a single word on-screen for 1200 ms, followed by a blank screen for 500 ms, and then the next word, until all 12 words were shown. A 10 s blank screen interval separated lists. Once all study blocks were completed there was a 5 min break which consisted of instructions for the Memory-guided saccade task and a short filler task. Afterwards participants underwent the memory-guided saccade task to test their memory of the studied word lists.

Memory-guided saccade task

Each block of the memory-guided saccade task was based on the previous 4 lists and consisted of 24 trials, with half of trials requiring a recognition decision (12 trials of seen list items, with 3 list words taken from position 1, 5, and 8 of each list), and half of trials requiring a non-recognition decision (8 trials of unrelated filler words, and 4 trials of critical lure items). Trial presentation was randomised. Each trial began with a fixation cross for 500–750 ms. The test word was then presented in the centre of the screen and at the same time a single 1° green target was shown (8° to the left or right of the word). All participants were instructed to make a prosaccade (i.e., make a saccade towards and fixate on the green target) if the word presented was one that they recognised from the study list just shown. However, if participants did not recognise the word then they were asked to make an antisaccade (i.e., make a saccade in the opposite direction of the green target). This means that for all studied words, the correct response was to make a saccade towards the target, whereas for unrelated and critical lures, the correct response was to make an antisaccade. The word and target remained visible for 2000 ms, followed by a 500 ms blank screen. This was repeated until all 24 trials in a block were completed, at which point participants were shown another 4 DRM lists and performed another block of the memory-guided saccade task, until all 6 blocks were completed. A single practice list with immediate test was provided at the start of the experiment so that participants could familiarise themselves with the method of response for recognition items. The participants were made aware that this was a word memory task and that when making their saccades, to do so as accurately and promptly as possible. The completion of all study-memory task blocks took 30 min.

Table 1. Saccade recognition rates and final corrected saccades (for Experiment 2 only) as function of item type across both experiments.

	Item Type		
	Studied words	Critical Lures	Unrelated Lures
Experiment 1 Recognition Rates: *M* (SD)	.71 (.10)	.77 (.13)	.11 (.08)
95% CI (LB, UB)	(.65, .77)	(.70, .85)	(.06, .15)
Experiment 2 Recognition Rates *M* (SD)	.72 (.09)	.76 (.14)	.07 (.08)
95% CI (LB, UB)	(.67, .76)	(.71, 82)	(.01, .12)
Final Corrected Saccades *M* (SD)	.74 (.10)	.78 (.15)	.03 (.05)
95% CI (LB, UB)	(.70, .78)	(.71, .84)	(.002, .05)

Note: *M*, SD, LB, and UB refer to Mean, Standard Deviation, Lower Bound, and Upper Bound for 95% confidence intervals (CI), respectively.

Results and discussion

To analyse recognition rates, we compared old recognition saccadic responses for critical lure, unrelated filler items and studied list items. Following a similar method of analysis as Coane et al. (2007), for saccadic latency, we wanted to analyse the efforts taken to make correct decisions (to either reject the critical lure and unrelated filler or accept the studied list item) as well as a direct comparison between list item hits and critical lure false alarms. Recognition rates and correct saccadic latencies were analysed using a one-way repeated measures analysis of variance ANOVA (word type: studied list item vs. unrelated fillers vs. critical lures). For the saccadic latency comparison of studied list item hits and critical lure false alarms, we used paired samples *t*-test. Bonferroni pairwise-comparisons (alpha set at .05) were used for comparisons across conditions. Mean proportions for recognition rates, including standard deviations, and 95% Confidence Intervals are reported in Table 1.

Recognition rates

For recognition rates, there was a significant difference in type of word, $F(2, 78) = 498.68$, $p < .001$, $\eta_p^2 = .93$. As expected, recognition rates for list items and critical lures were higher than unrelated fillers (both comparisons, $p < .001$). Recognition rates for critical lures were also higher than list items, $p = .004$.

Saccadic latencies

For correct saccadic latencies, there was a significant difference in word type, $F(2, 78) = 39.86$, $p < .001$, $\eta_p^2 = .51$. Pairwise comparisons revealed that latencies for correct identification of list items were faster than correct rejection of unrelated fillers ($p = .008$), but importantly, both were both faster than correct rejections of critical lure items (both $p < .001$).

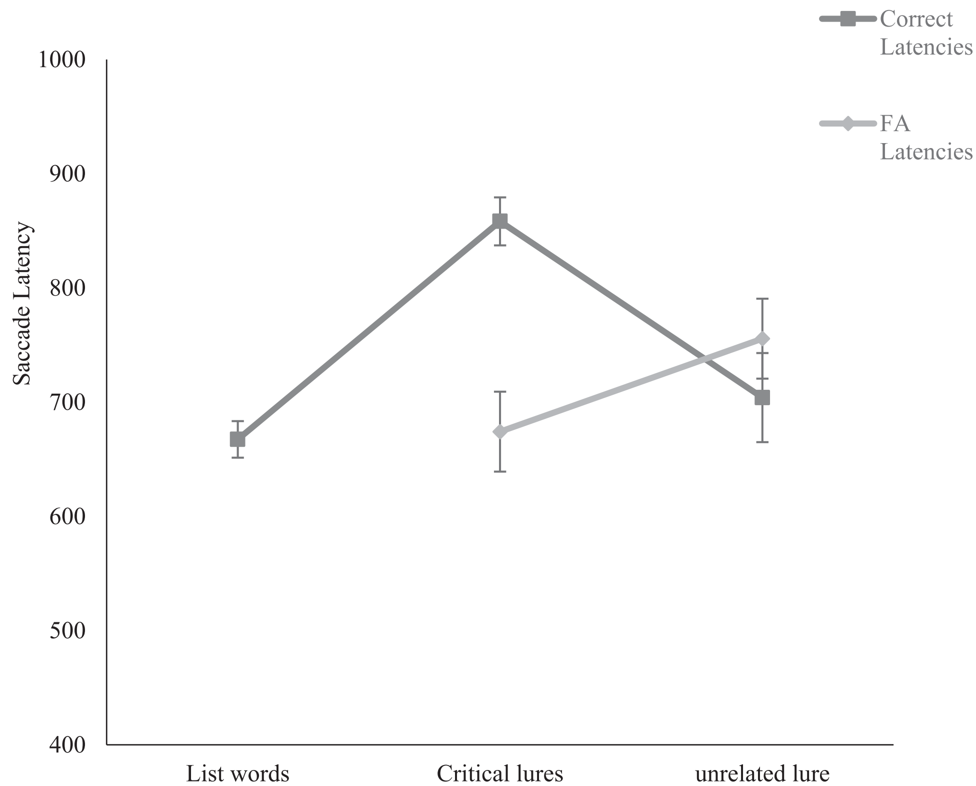

Figure 1. Correct (hits for list words and correct rejections for critical lures and unrelated lures) and FAs (incorrect false recognition of critical lures and unrelated lures) saccade latencies for Experiment 1. Error bars represent standard error.

This suggests that correct rejections of unrelated items and correct identifications of list items were easier to make than correct rejections of critical lures. In a separate analysis, we compared the saccadic latencies of old recognition decisions for list items (correct hit) and old recognition decisions for critical lures (false alarms to the critical lure). A paired-sample *t*-test revealed no statistical difference between the two latencies for these items, $t(39) = -1.70$, $p = .10$, $d = .14$ (see Figure 1).

When considering the findings from Experiment 1, first, it becomes evident that the saccade method effectively differentiates between recognition and non-recognition decisions, as evidenced by the minimal error rates for unrelated lures. For unrelated lures, the relative ease to reject (demonstrated by high non-recognition decisions and fast saccade latencies) is likely due to the low familiarity of the item providing sufficient evidence for cognitive processes to trigger the antisaccade. It is easier to make a correct rejection and thus inhibit the reflexive attention towards the target. This does not appear to be the case for critical lures. Our findings align with previous studies (e.g., Coane et al., 2007; Jou et al., 2004) showing that saccadic latencies for accurate rejections of critical lures were prolonged compared to those for unrelated lures. It takes additional effort to detect and engage source-monitoring to differentiate critical lure from list items, when such cognitive effort provides sufficient evidence, an antisaccade can be made. This will inevitably be slower.

Our final latency analysis compared list item hits versus critical lure false alarms to examine differences in saccade latency for true and false memories as measured by the DRM paradigm. Contrary to some response latency studies (e.g., Coane et al., 2007; Jou et al., 2004), but consistent with others (e.g., Thomas & Sommers, 2005; Tun et al., 1998), our data revealed no marked disparity between these two saccade latencies. We contend that, in contrast to the latency captured during explicit old/new responses, our saccade latency represents the ongoing cognitive processing. In our analysis, we interpret prosaccades as indicative of more reflexive, or automatic, decision-making processes. This is characterised by the eye's natural and rapid movement towards a visible target, known as fixation. Applying the activation–monitoring framework (Roediger & McDermott, 2000) to interpret our results, one could propose that the pronounced activation for the critical lure item exceeds the criterion threshold set for a faster *old* response when making a saccade decision, similar to that of a list item. The reflexive nature of the saccade response inhibits the opportunity to make a more controlled response, preventing any additional source-monitoring to occur before the decision is made. The speed of saccades may further support this suggestion. As shown in Table 1, average saccadic latencies were 674 ms for critical lure false alarms and 668 ms for list item hits, compared to approximately 1200 ms vs 1050 ms RTs respectively (values drawn from

Figure 1 in Coane et al., 2007). Such rapid saccades preempts the more deliberative phase of source-monitoring that typically aids in recognition. In Experiment 2 we aim to examine, if triggered, the point at which source-monitoring corrections may influence final recognition decision, something explicit responses and associated response latencies are unable to capture. We also changed our saccade task to remove any inherent response bias caused by fixating on a single peripheral target.

Experiment 2

In our second experiment, we adopted a two-target saccade task. The prosaccade in this task is no longer reflexive, it is rule based. There are, of course, methodological differences between standard DRM and saccade tasks aside from the modality of response that need to be addressed. The first is that stimulus-response compatibility (e.g., Proctor & Vu, 2006) is not a foremost issue in typical DRM response latency studies as response mapping for a recognition decision is always the same across trials (e.g., press x if recognised, y if not recognised). In contrast, the direction of the target (and therefore the correct response) in saccade tasks is always randomised. This demands participants to discern the correct response mapping for each trial. As such, the onset of a single peripheral target leads to a prepotent bias response. If eye movements are to be used to indicate a recognition decision, then this would bias responses towards the target location irrespective of the test word presented centrally (e.g., studied item, unrelated lure, critical lure). A two-target saccade task was designed to reduce responding bias to a single peripheral target, keeping fixed left and right targets to denote old/new recognition. This approach draws more similar parallels with traditional DRM response latency studies. Experiment 2 uses such a task by incorporating two distinctly coloured peripheral targets. This approach is more in line with the saccade tasks rooted in rule-based criteria as referenced in the introduction (Cherkasova et al., 2002; Rivaud-Pechoux et al., 2007). Participants received instructions to fixate on one of the designated colour targets (either blue or red) based on whether they recognised the centrally displayed test word, with each colour corresponding to recognition or non-recognition. These colour targets maintained a fixed position in each trial (consistently blue to the left and red to the right), though their positions were counterbalanced among participants.

For both experiments, our analysis encompasses both the saccade direction and the speed of the saccade to examine any such differentiation. In Experiment 2 we also examine corrective saccades, which is where an eye movement first made towards one target but is subsequently corrected, by making an eye movement in the opposite direction. Although any potential evidence of indecision or additional attentional resources can typically only be reflected in inflated RTs, we posit that these adjustments in decision-making are indicative of source-monitoring providing sufficient feedback that leads to the correction in the eye movement. Coupled with differences in saccadic latencies we hope to highlight if and when monitoring strategies occur during retrieval for true and false memories. The predictions for Experiment 1, still hold for Experiment 2, but additionally we predict that the nature and frequency of corrective saccades will vary depending on the type of word being recognised. For list items, which are true memories, we hypothesise that any initial incorrect recognition decisions (misses) will more likely be corrected by the end of the trial, leading to an increased rate of correct recognition (hits). This prediction is based on the assumption that true memories, once fully processed, are more readily recognised, even if initially missed. Conversely, for unrelated lures, which are entirely new and unfamiliar, we anticipate that participants will initially make some incorrect recognition decisions due to reflexive saccadic responses, but will often correct these decisions by the trial's end, resulting in a higher rate of correct rejections. However, for critical lures, we predict a different pattern. We expect that corrective saccades will be less effective in altering initial recognition decisions here. Participants may maintain their initial decision because it is more challenging to reassess the source of associative activation and sense of familiarity.

Method

Participants

Twenty participants (15 female) 18–57 years old ($M = 20.48$, $SD = 6.41$) from Edge Hill University, England were recruited for Experiment 2. Analysis from Experiment 1 indicated that a large effect size could be used. An a priori power analysis was therefore conducted using G*Power 3.1 to determine the required sample size that would be used for a repeated measures design with 3 conditions (our three item types). We set the effect size to $f = 0.4$. The power was set at ($\alpha = 0.05$, $1–\beta$ err prob) 0.95. We utilised a conservative correlation among repeated measures of 0.5. Based on these parameters, the power analysis indicated that a sample size of 18 participants would be necessary to achieve adequate power for detecting the expected effect. This sample size was considered sufficient for the repeated measures analysis of the corrective saccades. All participants had normal or corrected-to-normal vision and had English as their first language. They were rewarded with either course credit or monetary compensation.

Design

All participants completed the saccade task with the fixed target location design. Participants were presented with studied words, unrelated lures and critical lures during the saccade task. We examined the following dependent measures:

Recognition rate

This is whether participants looked at the appropriate target dot when presented with either the studied word, unrelated lure, or critical lures.

Saccadic latency (ms)

The time taken for the first eye movement towards the peripheral target since the test word was presented. This measure was identical to Experiment 1.

Corrected saccades

In addition to the above, we also wanted to examine whether the initial memory-guided decision was also the final response or whether participants changed their decisions in some way during the 2 s trial period. For example, whether an initial recognition decision of a critical lure was corrected to non-recognition decision in subsequent eye movements by participants looking at the opposite target. To do this we examined the final eye movement of each trial to examine whether it was located in the same location as the original eye movement or in the opposite spatial location. We define here a change in decision if the final eye movement was fully in the opposite direction to the first direction fixated – returning to the central fixation word position (–/+ 2°) would not count as a correction.

Stimuli and apparatus

The DRM lists were identical to those used in Experiment 1. A key difference in this experiment was that instead of only having a single green target to look towards (or away from) there were two targets (one red, one blue) that appeared simultaneously with the test word: 1 coloured target appeared to 8° to the left of the word, the other coloured target appeared to 8° to the right of the word. The location of the red and blue targets was fixed across all trials for the participant. This was counterbalanced so that for half of the participants, the red target was always to the left of the word, and the blue target was to the right.

Procedure

An information sheet and consent form were reviewed and signed before commencing with the study. As in Experiment 1, the study lists were presented in blocks of 4 and afterwards participant's memory was tested based on the 4 lists they had just seen. This resulted in 6 blocks in which participants undertook the memory-guided saccade task. The study phase was identical to Experiment 1.

Memory-guided saccade task

This followed a similar procedure except for the change in rules for making recognition and non-recognition decisions (old or new). Testing consisted of 6 blocks, with each block containing the previous 4 lists and consisted of 24 trials, half of trials requiring a recognition decision (12 trials of seen words, list items from position 1 and 8), and half of trials requiring a non-recognition decision (8 trials of unrelated filler words, and 4 trials of critical lure words). Word presentation was randomised and each trial began with a fixation cross for 500–750 ms. The test word was then presented in the centre of the screen and at the same time there were two target dots to the left/right of the word (one red, one blue). Participants were instructed to make a recognition decision by looking at the red target or a non-recognition decision by looking at the blue target (colour to indicate recognition was counterbalanced across participants), and the location of these colour targets was fixed throughout the experiment. A single practice list with immediate test was provided at the start of the experiment so that participants could familiarise themselves with the correct response for seen and unseen words. No participants failed to correctly use the red and blue target dots to make appropriate responses. The completion of all study-memory task blocks took 30 min to complete.

Results

To analyse recognition rates and correct saccadic latencies we used a repeated measures one-way ANOVA (Word Type: list items vs. unrelated lures, vs. critical lures). Again, for saccadic latency, we also conducted a separate analysis for the comparison of list item hits and critical lure false alarms. Finally, to assess whether corrected recognition decisions became "better" we compared recognition rates using a 3 (Word Type: Studied words vs. Unrelated lures vs. critical lures) × 2 (Response: Initial vs. Corrected) repeated measures ANOVA. Interactions were explored using simple main effects and Bonferroni pairwise-comparisons (alpha set at .05). Mean proportions for recognition rates and corrective saccades, including standard deviations, and 95% Confidence Intervals are reported in Table 1.

Recognition rates

There was a significant difference in word type, $F(2, 38) = 303.48$, $p < .001$, $\eta_p^2 = .94$. Bonferroni pairwise comparisons revealed that recognition rates for unrelated filler items significantly lower than for list items and critical lures (both $ps < .001$). For this experiment there was no statistical difference in recognition rates for list items and critical lures ($p = .45$).

Saccadic latencies

There were 2 missing cases for this latency analysis (2 participants failed to correctly reject any of the critical lures and so had no correct latencies for this word type).

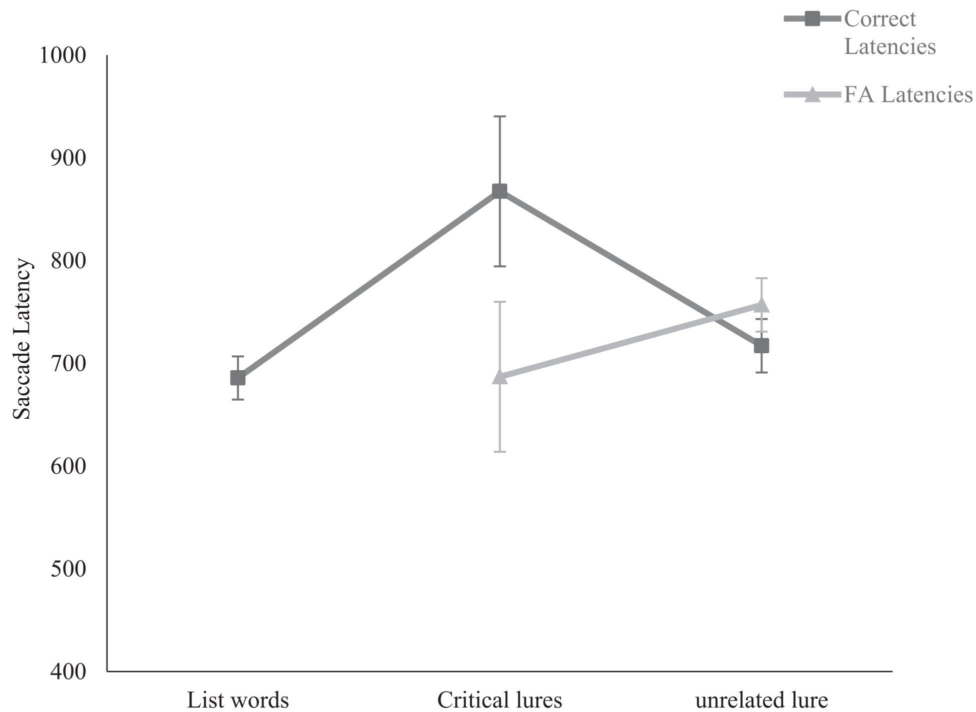

Figure 2. Correct (hits for list words and correct rejections for critical lures and unrelated lures) and FAs (incorrect false recognition of critical lures and unrelated lures) saccade latencies for Experiment 2. Error bars represent standard error.

Saccadic latencies for correct recognition decisions were significantly different across word type, $F(2, 34) = 7.28$, $p = .002$, $\eta_p^2 = .30$. Comparisons revealed that saccadic latencies were again faster for correct recognition of list items compared to correct rejection of critical lures ($p = .02$). Although the pattern of difference was in the same direction as Experiment 1, latencies for correct rejection of unrelated filler items was not significantly different to correct rejection of critical lures ($p = .092$), see Figure 2. Given the large effect reported in Experiment 1 we decided to use Bayesian techniques to determine the relative level of support for our hypothesis (that, as revealed by saccadic latencies, critical lures are harder to reject than unrelated filler items) over the null hypothesis. Bayes factors provide a continuous measure of how probable the data are under our hypotheses compared to how probable the data are under the null hypothesis. Software described in Dienes (2008, 2011, 2014) was used to undertake the Bayes factor calculation. This assumes, as a default, a null hypothesis where the true population value is equal to zero. The Bayesian approach demands specificity about the hypothesis to be contrasted with the null. Therefore, we assumed that the effect would vary in size between zero and the upper limit set by the effect size reported in Experiment 1. We based our prediction on a half-normal distribution wherein predicting smaller effect sizes is more likely than large effect sizes. Here, the estimate of the standard deviation of the p (population value|theory) was computed as the mean difference between the latencies of correct rejection of unrelated filler items and critical lures from Experiment 1 (154.28; SE = 25.93), and the mean of p (population value|theory) was set at 0. Data from Experiment 1 and 2 were combined in a meta-analytic manner using the mean and standard deviation (SD) from Experiment 1 as the prior mean and prior SD and the mean from Experiment 2 as the likelihood mean and likelihood's SD in order to calculate posterior mean and posterior SD. Once combined, a Bayes Factor was computed to represent the combined data from Experiments 1 and 2. The Bayes Factor was $BH(0, 154.28) = 1.48 \times 10^8$. Therefore, the results indicate extreme evidence for the alternative hypothesis over the null hypothesis (see Jeffreys, 1961).

There was also no significant difference between rejection of unrelated filler items and correct recognition of list items, $p = .31$. Similar to Experiment 1, when comparing saccadic latencies for correct recognition of list items and false recognition of critical lures, we found no significant difference, $t(19) = -.86$, $p = .40$, $d = .18$.

Corrective saccades

Overall, where changes were made, in some cases this meant that an initial incorrect decision was subsequently corrected by the end of the trial (e.g., initially recognising a critical or unrelated lure and making a saccade to the recognition target, but subsequently correcting this initial decision by the end of the trial by correctly fixating on the non-recognition target). However, the reverse was also true in some cases, whereby an initially correct recognition decision was subsequently changed to an incorrect decision by the end of the trial. To understand the net effect of these corrective saccades we analyse the initial recognition responses vs the final corrected responses for each word type. This analysis of corrective saccades revealed a significant main effect of word type, $F(2, 38) = 339.33$, $p < .001$, $\eta_p^2 = .95$. Although there was no main effect of response, $F(1, 19) = .01$, $p = .93$, $\eta_p^2 = .00$, there was a significant word type × response interaction, $F(2, 38) = 9.76$, $p < .001$, $\eta_p^2 = .34$. This interaction shows the change in response from initial to corrected recognition for each word type. Using separate paired samples *t*-tests, we found that the corrected final saccade increased correct recognition of list items, $t(19) = -2.78$, $p = .01$, $d = .21$ and decreased false alarms for unrelated filler items, $t(19) = 3.46$, $p = .003$, $d = .60$, however there was no significant change caused by corrective saccades for critical lures, $t(19) = -1.51$, $p = .15$, $d = .14$ (see Table 1 recognition rate and final corrective saccades). These results appear to suggest that whilst we are able to correct our decisions for the better when it comes to recognition of list items and unrelated lures, we do not appear to do this for recognition decisions for critical lures.

The results from Experiment 2 mainly replicated findings from Experiment 1, indicating that recognition decisions to critical lures and list items are made to a similarly high level compared to unrelated lures, and old responses to critical lures and list items are made at a similar speed (saccadic latency). Further, making a correct rejection of a critical lure is particularly demanding, taking longer than a false recognition of said critical lure and the correct recognition of a list item and the rejection of an unrelated lure (see Bayesian analysis supporting the latter). The novel finding from Experiment 2 relates to the corrective saccade analysis. By examining changes from initial to final saccade response, we are able to see the onset of source-monitoring used to trigger a correction. We observed that corrected saccades were in the positive direction for list items, that is, participants changed their initial response to increase correct recognition of list items moving from an incorrect miss to a correct hit. We also found positive gains for unrelated lures; corrective saccades showed that initial incorrect recognition decisions were subsequently changed to correct rejections. However, the corrective saccades for critical lures showed no move to correctly reject them from initial to final saccade, in fact, the pattern of corrections was for the worse, increasing false recognition. These findings are discussed further below.

General discussion

The two experiments presented here consider a novel paradigm to measure recognition and non-recognition decisions by using rule-based saccadic eye movements. The central focus of our investigation was to explore whether recognition rates, measured by saccadic eye movements, would reveal significant differentiation between the false recognition of critical lures and the accurate recognition of previously studied words. While recognition accuracy has been the predominant measure used to examine true and false memory differentiation in the Deese-Roediger-McDermott (DRM) literature, our research allowed for the examination of additional measures, such as saccadic latency and corrective saccades for true and false memory formation.

As expected, for recognition rates across the two experiments both critical lures and list items were recognised more often than unrelated lures. This replicates typical recognition findings reported when using the DRM paradigm (e.g., Roediger & McDermott, 1995). Interestingly, in the first experiment critical lures had a higher recognition rate than list items, although still similar saccadic latencies. For saccadic latencies, similar to Coane et al. (2007), we analysed two important comparisons. First, the saccadic latencies when making a correct recognition decision to accept a list item or reject a lure, and second, the comparison of saccadic latencies when making a false alarm to a critical lure and a correct hit to a list item. For correct recognition decisions, we saw that across the two experiments it took longer to make a correct rejection for a critical lure compared to the correct recognition of a list item. Experiment 1 also found that the correct rejection of an unrelated filler was faster compared to the correct rejection of the critical lure. Experiment 2 showed a similar pattern, and although the frequentist statistic was not significant, $p = .09$, our use of Bayesian techniques to combine data from Experiments 1 and 2 yielded extreme evidence for the alternative hypothesis that correct rejections for critical lures were slower than for unrelated fillers. These latter findings support previous response latency research (Coane et al., 2007; Jou et al., 2004) and the notion that critical lures require additional processing to monitor and reject the item. If we consider activation–monitoring theories (Roediger et al., 2001; Roediger & McDermott, 1995, 2000), slower correct rejections for critical lures are likely a result of strength of activation not being sufficient to prevent the onset of additional differentiating evidence controlling the saccade. Decision mechanisms can use relatively undifferentiated information for old–new recognition decisions but require more differentiated information for source-monitoring decisions. It takes time for the information that makes up a memory to become differentiated (Johnson et al., 1993). Consistent with this idea is the finding that antisaccadic latencies or non-recognition latencies are slower for the correct rejection of the critical lure. This encompasses the time taken to detect the discrepancy and activate source-monitoring. For unrelated lures, the cognitive process is more straightforward: since these items do not trigger any memory recall or recognition, one can quickly conclude that they are new or irrelevant. This rapid assessment allows for a swift initiation of the antisaccade or non-recognition response. There is no need for extended cognitive evaluation or "evidence accumulation" as the item is immediately identified as unfamiliar. Consequently, the participant can promptly initiate the appropriate eye movement or recognition response without the delay seen in responses to critical lures.

The direct comparison of false alarm and list item saccadic latencies provides a secondary true and false memory measure. The findings in the response latency literature are mixed when it comes to differentiation between the two, with some studies showing no response latency difference (Hancock et al., 2003, specifically Experiments 3 and 4; Thomas & Sommers, 2005; Tun et al., 1998 [older adults only]), while others have shown false alarms to critical lures to be consistently slower than correct hits (Coane et al., 2007; Jou et al., 2004; Jou et al., 2017). In both experiments, we found no significant difference in saccadic latencies between true and false memories. We have seen evidence that slower correct rejections for critical lures demonstrate associative activation and monitoring processes acting in opposition. With automatic activation of the semantic associates, familiarity for these critical lures exceeds the criterion for accepting a word as part of the study list, before source-monitoring processes have time to detect a discrepancy (see Coane et al., 2007 explanation). Although these findings are not consistent with those of Jou et al. (2004, 2017) or Coane et al., they could perhaps be explained by the faster saccade latencies we see here compared to the response latencies reported in previous studies. All previous response latencies have been much slower than ours. Coane et al., report mean response times for false alarms to critical lures as approximately 1200–1300 ms (even Tun et al, Thomas & Sommers approx. mean response latencies were 1000–1200 ms). Our mean latency time was 674.18–686.86 ms. Keyboard responses are inherently slower and may still invoke time for source-monitoring, although ultimately being accepted as old. Coane et al., themselves suggested future studies might investigate the effects of shorter response deadlines (e.g., 500 ms) that would require subjects to rely primarily on familiarity responses. Indeed, speeded response studies that allow only 800 ms for a response see greater false memories compared to self-paced recognition decision (Carneiro et al., 2017). We propose that our response latencies make invoking controlled monitoring processes all the more harder when latencies are very fast and associative activation triggers a high level of familiarity. In addition, very recent research (Kafkas et al., 2023) has used pupil response patterns to distinguish true from false memories in the DRM paradigm. They too suggest that early pupil dilation was related to false familiarity responses to the critical lure, while later pupil dilation would suggest recovery of source and contextual details for recollection. Pupil response appears to provide additional retrieval-based information during true and false memories. Although we do note that prior pupil response data has shown mixed findings (Gomes et al., 2021; Montefinese et al., 2013; Otero et al., 2011), and there is some debate as to whether pupil output is linked to other factors, such as working memory load, arousal or response preparation. Nevertheless, we see our work as complementary to these pupil approaches, both providing additional information compared to the more traditional keyed input response.

One final measure that we investigated in Experiment 2 was corrective saccades. We have argued that because initial saccades are made much faster on average than typical keyed response latencies, they could represent a more automatic decision, before a controlled process of source-monitoring can aid recognition decisions (Coane et al., 2007). In the final experiment, we therefore examined whether participants would make any corrective saccades after their initial saccade decision. We analysed changes between initial and final saccade for list items as well as critical and unrelated lures. If we consider models of saccade generation (see Boucher et al., 2007 for a review), we argue that any corrections represent source-monitoring processes providing sufficient evidence to confirm the validity of the attribution which could thus trigger a change in saccade. We found that corrective saccades improved correct recognition for list items, moving from an initial incorrect miss to a correct hit. The initial miss was corrected by the onset of source-monitoring. Results also showed that participants could correct initial false alarms of unrelated lures. These initial false alarms may have been a result of cognitive bias or low threshold false familiarity prior to the onset of source attribution-based recollection. When the false familiarity threshold is low enough to detect this error and trigger the change. Surprisingly, however, when participants made a false alarm response to a critical lure, we saw no significant corrections, that is, participants did not correct their initial decision. In the introduction, we argued that saccades will be based on faster old/new recognition decisions based on familiarity. This can be seen in their rapid execution. We argued that corrections may represent a more controlled decision to override that response. This is easy when the level of activation is low. For unrelated lures a decision can be made to override the initial response and reject the item because the item fails to elicit recollections that are expected from prior occurrence (Gallo, 2010). They can correct that reflexive response with a more control decision process. This does not appear to happen for critical lures when they are initially categorised as old. Saccades are fast compared to the end result of a response time. They are more implicit and less easily controlled. It takes more effort to stop the prosaccade recognition of the critical lure. Here it is plausible to argue that false alarms for critical lures are likely not corrected in this saccade task because of the relatively automatic process by which the semantic activation occurs (Underwood's, 1965 implicit-associative-response [IAR] theory) and how such items, when generated, are treated as list members. Once recognised, the threshold is sufficient that it passes some expected criteria to prevent additional source-monitoring checking processes. Although we have aligned these explanations with spreading activation and source monitoring frameworks, it also resonates with fuzzy-trace perspectives (Brainerd & Reyna, 2002). In this perspective, the strong sense of familiarity for critical lures can be attributed to gist extraction. Gist processing, which captures the theme of words, facilitates the perception of critical lures as previously encoded events. This process is typically automatic and requires less cognitive effort. Conversely, the recovery and validation of source and contextual details, crucial for accurate recognition, align with verbatim processing. The formation and retrieval of verbatim traces require more cognitive resources and time, as they involve recalling specific details and features. It's important to note that our findings do not favour one theoretical explanation over another, but the source-monitoring component of the activation–monitoring framework aligns well with the corrective saccades that we analysed in Experiment 2.

To conclude, a great advantage of saccadic tasks, one which is only just beginning to be exploited by researchers interested in cognitive processes, is that they allow the ongoing control of behaviour to be studied (Hutton, 2008). We believe that the saccadic task used here to measure recognition accuracy and response latency, and corrective decisions, adds to the literature examining differentiation of the true and false memories in the DRM paradigm. By using an eye-tracker to record eye movement behaviour, it is possible to reveal rapid visual and cognitive processing in a given task. To our knowledge, this memory-guided manipulation has not been implemented before to study false memories. The DRM-Saccade task certainly provides a novel approach to measuring cognitive processes in recognition decisions and we hope that others may adopt this methodology. We say this because standard recognition memory paradigms allow researchers to only examine the end product of memories, the actual decision, response time, and the subjective memory strength. The DRM-saccade task presented here has the potential to demonstrate memory processes at work and indexes cognitive resources during retrieval. While the experiments reported here offer a proof of concept, the future use of the DRM-saccade task could further help understand the role source-monitoring and its onset during recognition decisions. Whilst by no means, an exhaustive list, it could be used to assess individual differences and cognitive decline in relation to source monitoring, speed of processing and the ability to use this information to make real time corrections. Research highlights various experimental manipulations to examine the role of source-monitoring and source attribution, and the DRM-saccade task could provide additional retrieval information when utilising these manipulations. Furthermore, integrating the DRM-Saccade task with neuroimaging techniques or physiological measures, such as EEG or fMRI, could facilitate a multi-modal exploration of the neural correlates underlying memory processing. The DRM-Saccade task, thus, stands as a promising avenue for future research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data that support the findings of this study are available from the corresponding author, Lauren Knott, upon request.

References

Boucher, L., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2007). Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review*, *114*(2), 376–397. https://doi.org/10.1037/0033-295X.114.2.376

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Brown, N. R., Buchanan, L., & Cabeza, R. (2000). Estimating the frequency of nonevents: The role of recollection failure in false recognition. *Psychonomic Bulletin & Review*, *7*(4), 684–691. https://doi.org/10.3758/BF03213007

Carneiro, P., Garcia-Marques, L., Lapa, A., & Fernandez, A. (2017). Explaining the persistence of false memories: A proposal based on associative activation and thematic extraction. *Memory*, *25*(8), 986–998. https://doi.org/10.1080/09658211.2016.1239742

Cherkasova, M. V., Manoach, D. S., Intriligator, J. M., & Barton, J. J. (2002). Antisaccades and task-switching: Interactions in controlled processing. *Experimental Brain Research*, *144*(4), 528–537. https://doi.org/10.1007/s00221-002-1075-z

Coane, J. H., McBride, D. M., Raulerson, B. A., & Jordan, J. S. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*(1), 62–70. https://doi.org/10.1027/1618-3169.54.1.62

Crawford, T. J., Higham, S., Renvoize, T., Patel, J., Dale, M., Suriya, A., & Tetley, S. (2005). Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease. *Biological Psychiatry*, *57*(9), 1052–1060. https://doi.org/10.1016/j.biopsych.2005.01.017

Cutsuridis, V., Smyrnis, N., Evdokimidis, I., & Perantonis, S. (2007). A neural model of decision-making by the superior colicullus in an antisaccade task. *Neural Networks*, *20*(6), 690–704. https://doi.org/10.1016/j.neunet.2007.01.004

Deese, J. (1959). On the prediction of occurrence of verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fischer, B. (1987). The preparation of visually guided saccades. In *Reviews of Physiology, Biochemistry and Pharmacology* (Vol. *106*, pp. 1–35). Berlin, Heidelberg: Springer. https://doi.org/10.1007/BFb0027574

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory and Cognition*, *38*(7), 833–848. https://doi.org/10.3758/MC.38.7.833

Gomes, C. A., Montaldi, D., & Mayes, A. (2021). Can pupillometry distinguish accurate from inaccurate familiarity? *Psychophysiology*, *58*(8), e13825. https://doi.org/10.1111/psyp.13825

Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation level of critical lures in the Deese-Roediger-McDermott paradigm. *The American Journal of Psychology*, *116*(1), 1–14. https://doi.org/10.2307/1423332

Hanes, D. P., & Carpenter, R. H. S. (1999). Countermanding saccades in humans. *Vision Research*, *39*(16), 2777–2791. https://doi.org/10.1016/S0042-6989(99)00011-5

Hanes, D. P., & Schall, J. D. (1995). Countermanding saccades in macaques. *Visual Neuroscience*, *12*(5), 929–937. https://doi.org/10.1017/S0952523800009482

Hutton, S. B. (2008). Cognitive control of saccadic eye movements. *Brain and Cognition*, *68*(3), 327–340. https://doi.org/10.1016/j.bandc.2008.08.021

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

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

Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1409–1419. https://doi.org/10.1037/0278-7393.20.6.1409

Jou, J., Arredondo, M. L., Li, C., Escamilla, E. E., & Zuniga, R. (2017). The effects of increasing semantic-associate list length on the Deese–Roediger–McDermott false recognition memory: Dual false-memory process in retrieval from sub- and supraspan lists. *The Quarterly Journal of Experimental Psychology*, *70*(10), 2076–2093. https://doi.org/10.1080/17470218.2016.1222446

Jou, J., & Foreman, J. (2007). Transfer of learning in avoiding false memory: The roles of warning, immediate feedback, and incentive. *The Quarterly Journal of Experimental Psychology*, *60*(6), 877–896. https://doi.org/10.1080/17470210600831184

Jou, J., Matus, Y. E., Aldridge, J. W., Rogers, D. M., & Zimmerman, R. L. (2004). How similar is false recognition to veridical recognition objectively and subjectively? *Memory & Cognition*, *32*(5), 824–840. https://doi.org/10.3758/BF03195872

Kafkas, A., Brown, T., Olusola, N., & Guo, C. (2023). Pupil response patterns distinguish true from false memories. *Scientific Reports*, *13*(1), 17244. https://doi.org/10.1038/s41598-023-44362-6

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review*, *6*(1), 130–135. https://doi.org/10.3758/BF03210820

Lampinen, J. M., Ryals, D. B., & Smith, K. (2008). Compelling untruths: The effect of retention interval on content borrowing and vivid false memories. *Memory*, *16*(2), 149–156. https://doi.org/10.1080/09658210701839277

Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: A model and a method. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(2), 276–291. https://doi.org/10.1037/0096-1523.10.2.276

Lyle, K. B., & Johnson, M. K. (2006). Importing perceived features into false memories. *Memory*, *14*(2), 197–213. https://doi.org/10.1080/09658210544000060

McDermott, K. B., Gilmore, A. W., Nelson, S. M., Watson, J. M., & Ojemann, J. G. (2017). The parietal memory network activates similarly for true and associative false recognition elicited via the DRM procedure. *Cortex*, *87*, 96–107. https://doi.org/10.1016/j.cortex.2016.09.008

Miller, A. R., Baratta, C., Wynveen, C., & Rosenfeld, J. P. (2001). P300 latency, but not amplitude or topography, distinguishes between true and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 354–361. https://doi.org/10.1037/0278-7393.27.2.354

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The "subjective" pupil old/new effect: Is the truth plain to see? *International Journal of Psychophysiology*, *89*(1), 48–56. https://doi.org/10.1016/j.ijpsycho.2013.05.001

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218–228. https://doi.org/10.1038/nrn1345

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida Word Association, rhyme and fragment norms* [Data file]. University of South Florida Web Site. http://luna.cas.usf.edu/nelson/

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory*, *9*(1), 53–71. https://doi.org/10.1080/09658210042000076

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*(6), 838–848. https://doi.org/10.3758/BF03211328

Osman, A., Kornblum, S., & Meyer, D. E. (1986). The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 243–258. https://doi.org/10.1037/0096-1523.12.3.243

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *48*(10), 1346–1353. https://doi.org/10.1111/j.1469-8986.2011.01217.x

Proctor, R. W., & Vu, K.-P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application*. CRC Press.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rivaud-Pechoux, S., Vidailhet, M., Brandel, J. P., & Gaymard, B. (2007). Mixing pro- and antisaccades in patients with parkinsonian syndromes. *Brain*, *130*(1), 256–264. https://doi.org/10.1093/brain/awl315

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology*, *21*(4), 803–814.

Roediger, H. L., III, & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, *9*(4), 123–127. https://doi.org/10.1111/1467-8721.00075

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*(3), 385–407. https://doi.org/10.3758/BF03196177

Schacter, D. L., Chamberlain, J., Gaesser, B., & Gerlach, K. D. (2012). Neuroimaging of true, false, and imaginary memories: Findings and implications. In L. Nadel & W. P. Sinnott-Armstrong (Eds.), *Memory and law* (pp. 233–262). Oxford University Press.

Schacter, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron*, *44*(1), 149–160. https://doi.org/10.1016/j.neuron.2004.08.017

Seideman, J. A., Stanford, T. R., & Salinas, E. (2018). Saccade metrics reflect decision-making dynamics during urgent choices. *Nature Communications*, *9*(1), 2907. https://doi.org/10.1038/s41467-018-05319-w

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language*, *52*(1), 71–86. https://doi.org/10.1016/j.jml.2004.08.001

Trappenberg, T. P., Dorris, M. C., Munoz, D. P., & Klein, R. M. (2001). A model of saccadic initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *Journal of Cognitive Neuroscience*, *13*(2), 256–271. https://doi.org/10.1162/089892901564306

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology & Aging*, *13*(2), 230–241. https://doi.org/10.1037/0882-7974.13.2.230

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*(1), 122–129. https://doi.org/10.1037/h0022014

Wilimzig, C., Schneider, S., & Schöner, G. (2006). The time course of saccadic decision making: Dynamic field theory. *Neural Networks*, *19*(8), 1059–1074. https://doi.org/10.1016/j.neunet.2006.03.003

Appendix. 24 DRM word lists (critical lures in bold) used for both Experiments 1 and 2

car	needle	chair	sleep	rough	mountain	soft	high
truck	thread	table	bed	smooth	Hill	hard	low
bus	pin	sit	rest	sandpaper	peak	velvet	up
vehicle	eye	legs	wake	tough	climber	loud	tall
bike	sewing	seat	tired	bumpy	hike	fluffy	tower
drive	sharp	couch	dream	rigid	valley	tender	jump
jeep	point	desk	yawn	gentle	summit	gentle	above
garage	prick	recliner	snooze	harsh	climb	cotton	building
van	thimble	sofa	blanket	course	slope	fur	noon
taxi	haystack	bench	doze	crude	rocks	silk	cliff
train	thorn	cushion	slumber	grit	steep	touch	sky
race	injection	sitting	lazy	uneven	canyon	feather	dive
caravan	knitting	stool	peace	surface	cave	skin	elevate
doctor	**cat**	**sweet**	**smoke**	**foot**	**thief**	**slow**	**man**
nurse	kitten	sour	cigar	shoe	robber	fast	woman
sick	dog	candy	cigarette	hand	crook	lethargic	husband
medicine	mouse	sugar	puff	toe	burglar	snail	uncle
health	pounce	bitter	chimney	kick	stolen	turtle	lady
hospital	fluffy	taste	tobacco	sandals	robbery	quick	mister
dentist	claw	tooth	pipe	walk	bandit	sluggish	male
physician	whiskers	honey	lungs	ankle	theft	lazy	father
patient	tiger	chocolate	inhale	boot	criminal	stall	human
stethoscope	pet	cake	pollution	inch	steal	delay	person
surgeon	tail	tart	flames	yard	beggar	hurry	handsome
clinic	sphinx	tangy	blaze	sock	liar	hesitate	male
ill	paw	dessert	ashes	knee	convict	cautious	boss
cold	**smell**	**cup**	**window**	**music**	**river**	**shirt**	**pen**
hot	nose	saucer	door	Band	stream	blouse	pencil
chill	sniff	mug	glass	concert	flow	sleeves	write
arctic	aroma	measuring	pane	violin	water	pants	fountain
ice	hear	glass	shade	stereo	creek	tie	quill
winter	see	measure	ledge	tune	bridge	button	felt
frost	whiff	coaster	sill	radio	brook	shorts	Bic
snow	scent	handle	open	jazz	lake	polo	scribble
wet	reek	coffee	curtain	piano	canal	collar	crayon
weather	stench	drink	frame	song	raft	vest	marker
freeze	fragrance	plastic	view	record	channel	pocket	cap
warm	perfume	lid	screen	album	waterfall	belt	point
frozen	sense	sip	shutter	volume	rapid	cuffs	blot