Source Reinstatement in Item Method Directed Forgetting Influences Recognition Strategies

Kathleen L. Hourihan

Department of Psychology, Memorial University of Newfoundland

Author Note

Kathleen L. Hourihan https://orcid.org/0000-0002-1087-9915

I have no conflicts of interest to declare. This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant 418242-2013. I thank Madison Gregory for her assistance with data collection.

Correspondence concerning this article should be addressed to Kathleen L. Hourihan, Department of Psychology, Memorial University of Newfoundland, St. John’s, NL, A1B 3X9, CANADA. Phone: (709) 864-8771. Email: khourihan@mun.ca

Abstract

In item method directed forgetting, participants study items paired with instructions to either remember or forget each item for the purpose of an upcoming memory test. Such instructions are effective, in that participants recall or recognize more remember- than forget-cued items when asked to disregard the cues at test. Recent research has shown that context and source information associated with targets at encoding are not subject to any influence of directed forgetting, such that both remember and forget items benefit equivalently from context reinstatement at test. In the current study, remember and forget items were presented by two sources, one of which presented mostly-remember items and one of which presented mostly-forget items. When the sources were reinstated at recognition, participants displayed more liberal responding to the mostly-remember source, such that item discriminability was actually worse compared to the mostly-forget source. When source information is reinstated at test, participants use their knowledge about the sources heuristically when making recognition judgments.

Keywords: directed forgetting, source memory, source reinstatement, response criterion

Public Significance Statement

Participants studied words presented by two different sources: one source provided mostly important information and the other source provided mostly unimportant information. When memory for the information was tested, results showed that participants were more cautious when judging their memory for information from the source who provided mostly unimportant information. We change the manner in which we assess our own memory, depending on the reliability of the source of that information.

Source Reinstatement in Item Method Directed Forgetting Influences Recognition Strategies

In an information-rich environment, the ability to selectively encode relevant material—and to disregard irrelevant material—is an adaptive cognitive ability. In the item method directed forgetting (DF) paradigm, participants are provided with a list of items (normally words) to learn for an upcoming memory test, but each item is followed by a cue to either Remember (R) or Forget (F) that word. When participants try to recall or recognize all studied items, regardless of cue, a DF effect is nearly always observed: memory for R items exceeds memory for F items (see MacLeod, 1998 for a review). Most explanations of the item method DF effect are based on a selective rehearsal account. That is, participants are aware that each study item will be followed by a memory cue; some items will be irrelevant for the upcoming test, and will receive an F cue. In typical DF procedures, the target item is removed from the screen before the cue is presented, and therefore each item is believed to be maintained in working memory in a shallow manner (e.g., via rote rehearsal) only until the cue is presented. If an R cue is presented, then participants will initiate deeper, more elaborative (meaningful) rehearsal (see Craik & Lockhart, 1972), but if an F cue is presented, then the item is removed from the rehearsal set (although exactly how F items are removed from rehearsal is under debate; e.g., Bancroft et al., 2013; Fawcett & Taylor, 2008; Zacks et al., 1996). In the current study, I examine how recognition decisions about R and F cued items are influenced by source information.

The item method DF effect is quite robust concerning memory for the target item (e.g., Burgess et al., 2017; Hourihan, 2021; Hourihan & Fraundorf, 2021; MacLeod, 1999; Taylor & Hamm, 2018; Thompson et al., 2011). However, it is only relatively recently that researchers have begun to examine how memory for context or source information (see, e.g., Johnson et al., 1993) is influenced in the item method DF paradigm. Perhaps unsurprisingly, the foundations of this recent interest in how DF influences context memory come from Dr. Bill Hockley. Dr. Hockley's research has informed us on item and associative recognition over the past several decades (e.g., Hockley, 1992; 2008; Hockley & Consoli, 1999; Hockley & Cristi, 1996). Indeed, he has provided unique perspectives on forgetting of associative information, both unintentionally (e.g., Hockley, 1991; 1992) and intentionally (Hockley et al., 2016).

Following my 2015 conference presentation of some initial data on context memory in item method DF, Bill approached me to discuss some of his own initial data on the topic, and proposed that we collaborate. I was delighted by the opportunity to work with such a renowned researcher, and we co-supervised Nicole Burgess's honours thesis research on context in item method DF. In three experiments, participants studied R and F cued words that were presented on unique background images as context. Burgess et al. (2017) showed that although both R and F items benefitted from context reinstatement at recognition, they did so equivalently, such that memory for context does not appear to be influenced by selective rehearsal. Taylor and Hamm (2018) showed a similar result using a different manipulation of context.

Hourihan and Fraundorf (2021) examined source memory in item method DF by presenting study items paired with one of two face icons representing different sources. Each source presented an equal number of R and F cued words. At test, participants first provided an old/new recognition judgment, and then provided a source judgment following "old" responses. They found a typical DF effect in item recognition, and further showed that source memory (i.e., source attribution given correct recognition) was equivalent for recognized R and F words. Following Hourihan and Fraundorf (2021), Hourihan (2021) presented participants with words paired with one of two sources in an item method DF paradigm. Importantly, the two sources presented R and F cues with different probabilities, such that one source (mostly-remember) presented twice as many R cues as F cues, and the other source (mostly-forget) presented twice as many F cues as R cues. Item and source memory were tested similarly to Hourihan and Fraundorf (2021). These experiments showed that the different cue probabilities influenced how participants rehearsed selectively, resulting in higher hit rates to words (F cued words in particular) from the mostly-remember source than from the mostly-forget source. Analysis of overt rehearsal showed that words from the mostly-remember source were more likely to be rehearsed prior to the presentation of the memory cue than words from the mostly-forget source. That is, F cued words were more likely to be recognized when presented by the mostly-remember source because participants' encoding strategies were influenced by knowledge of the cue probability; this knowledge led to differential rehearsal strategies depending both on cue and source. Moreover, multidimensional processing tree (MPT) modelling of source judgments (Bayen et al., 1996) showed differential source guessing depending on the memory cue: R cued words were more likely to be guessed as having come from the mostly-remember source, and F cued words were more likely to be guessed as having come from the mostly-forget source.

Thus, there is clear evidence that the cue probabilities associated with different sources can influence encoding processes. In the current paper, I use a paradigm similar to Hourihan (2021), but with reinstatement of the study sources at test to examine the influence this has on participants' recognition decisions. That is, rather than asking participants to provide a source judgment after labelling an item as "old" at recognition, the sources were re-presented at test along with the words when recognition judgments were elicited. Each test source presented R and F words that had been previously studied with that source (i.e., the correct source was reinstated), and an equal number of new words (see Table 1). Given my earlier finding that participants were more likely to guess the mostly-remember source for R cued words when making a source judgment, it follows that participants may use a similar heuristic when making a judgment in the opposite direction: guessing that a word was R cued if they knew that it came from the mostly-remember source. Thus, reinstating the sources at test has significant potential to influence recognition response decisions about whether an item was studied, and with which memory cue, based on participants' expectations about source-cue contingencies.

Past research has shown that memory expectations at test can influence decision strategies. For example, the distinctiveness heuristic can be used to reduce false recognition when items are studied in a distinctive manner (such as being read aloud; e.g., Dodson & Schacter, 2001). McDonough and Gallo (2012) showed that expectations about memory for certain items can influence recognition test judgments, even when there is no actual difference in memory strength. In their work, participants expected certain studied items (those presented in larger font) to be better remembered than others (presented in smaller font), and consequently used more conservative response criteria at recognition, reducing false recognition of familiar lures. This was true even though the memory differences were illusory, based on a false belief that large font words would be remembered better than small font words (see Rhodes & Castel, 2008); font size did not actually affect memory strength. Once again, Bill Hockley's research contributions are relevant to my current research, as he has shown that participants can adopt different response criteria on a per-trial basis (Hockley & Niewiadomski, 2007) in response to item strength.

Hourihan (2021) also provided evidence that participants' expectations about source-cue contingencies affected decision strategies at test. In particular, Experiment 2 elicited cue-tagging judgments at test, wherein participants responded whether they believed an item had been R-cued, F-cued, or was New (see Thompson et al., 2011). For "R" and "F" responses, participants then judged whether the item had been presented by the mostly-remember or the mostly-forget source. Results showed that participants made source judgments that were linked to their reported cue tag, rather than the actual cue. That is, regardless of whether an item had been presented with an R or F cue, if participants called an item "R", they were more likely to also choose the mostly-remember source, and if an item was called "F", they were more likely to choose the mostly-forget source.

Throughout the four experiments in Hourihan (2021), item recognition hit rates were generally higher for the mostly-remember source than for the mostly-forget source. The pattern of hit rates, along with the evidence from overt rehearsal, indicates that encoding processes were influenced by participants' knowledge about source cue probability. Of course, retrieval processes also have the potential to be influenced by participants' strategies. However, with only one type of new item at test, it was not possible to examine whether participants used different response criteria at test, depending on their belief about the source. Therefore, in the current study, the goal was to examine whether recognition decisions were influenced by knowledge about source cue probabilities. To do this, participants were not asked to provide a separate source judgment at test; instead, each test trial presented one of the same two sources as at study along with the to-be-tagged word. Each source presented its originally studied words (i.e., the correct source was always reinstated for studied items) along with an equal number of new words, and participants were asked to provide an R-F-N tagging response for each item. This procedure allowed for calculation of separate false alarm rates for the two test sources to determine whether participants adopt different criteria based on the likelihood that they had intentionally studied information from a given source.

Thus, examining whether participants use different response criteria for the two test sources is informative for how source knowledge influences the retrieval aspect of memory in this paradigm. It was expected that participants may adopt a more conservative criterion for recognizing items from the mostly-forget test source. In the present experiment, rather than relying on belief or making a guess based on memory strength (as in Hourihan, 2021), the original source is re-presented at the time of the recognition tagging decision. Participants know that one source, the mostly-forget source, primarily presented words that were F cued, and were therefore unlikely to have been rehearsed elaboratively. This source, however, did still present some R cued items that should have been rehearsed. Overall, recognition decisions about items from this test source should be evaluated relatively cautiously, as participants may expect that most studied items would be relatively low in strength, and may be difficult to differentiate from new, unstudied items. Conversely, the mostly-remember source primarily presented R cued words that were rehearsed, but also a smaller number of words that were F cued (which may still have received more pre-cue rehearsal than words from the mostly-forget source;

Hourihan, 2021). If participants assume that most words from this test source were rehearsed, then they may make recognition judgments in accordance with this expectation, and be less cautious about evaluating memory strength than for the mostly-forget test source. It was therefore expected that false alarms would be higher for new items from the mostly-remember test source than from the mostly-forget test source, and response criterion (here, *c*) would be more liberal for the mostly-remember test source.

In addition to comparing response criteria between the two test sources, hits and false alarms were used to compute *d'*, a signal detection-based measure of discriminability (suitable for old/new judgments). I predicted that overall hit rate would follow the same pattern as in Hourihan (2021): higher hits for remember than forget words, and potentially higher hits for the mostly-remember source than for the mostly-forget source (although this source effect may only be apparent for F words). When item discriminability is computed taking into account response criterion differences, the effect of source cue probability may be eliminated. The effect of memory cue in hit rates and discriminability was expected to be significant, as the DF effect is likely much more reliant on actual memory strength differences caused by selective rehearsal than on recognition decision strategies (see also Taylor et al., 2018). Tagging accuracy, however, was expected to be influenced by knowledge about a test source's study phase cue probability in a way similar to the source memory judgments found by Hourihan (2021). Here, the mostly-remember test source was therefore expected to lead participants to provide more R tags than F tags, and the mostly-forget test source to lead participants to provide more F tags than R tags. This would be consistent with participants' expectations about which sorts of cues were more likely to have been presented by the two sources when they had been encountered in the study phase.

Method

Participants

Thirty-four undergraduate students at Memorial University of Newfoundland participated in person in exchange for either course credit or $10 payment. The target number of participants planned was 32, based on previous studies examining source or context memory in item method DF (e.g., Burgess et al, 2017; Hourihan, 2021; Hourihan & Fraundorf, 2021; Thompson et al., 2011). Two additional participants were tested due to the nature of the credit participation recruitment process; data from one participant were lost due to computer error resulting in a final sample of 33 participants.

Materials

The same materials were used as in Hourihan (2021). A list of 200 words was generated from the MRC psycholinguistic database (Coltheart, 1981). Words were between five and nine letters in length, with a mean frequency (Kucera & Francis, 1967) of 48.8 [5, 198]. Words were randomly selected from this pool and assigned to condition for each participant.

A female and male face icon were used as the two sources, obtained from freedigitalphotos.net. The icons depicted an outline of a face, with different hairstyles and background colours for the two sources (see Figure 1). The assignment of face icons to source cue probability condition was randomly determined for each participant. E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used to present the stimuli and record all responses. The study was reviewed and approved by the Memorial University Interdisciplinary Committee on Ethics in Human Research (#20160772).

Procedure

The study phase of the experiment was identical to that used by Hourihan (2021); see Figure 1 for example trials. Participants were informed that they would be given a list of words to learn for a memory test, but that not all items would appear on the test; words followed by “RRRRR” would be on the test and should be remembered while words followed by “FFFFF” would not be on the test, and could be forgotten. A second instruction screen displayed the two face icons, one on the left and one on the right, and participants were told that words would be given to them by two different people. They were further told that one of the sources would give them mostly F words that would not be tested and the other source would give them mostly R words that would be tested. Assignment of cue probability condition to the source on the left or right side of the screen was counterbalanced across participants, based on order of appearance to the lab. Following four buffer trials (two from each source, with cues consistent with their stated cue probability), 120 study trials were presented in random order (See Table 1). The mostly-remember source presented 40 R and 20 F words; the mostly-forget source presented 20 R and 40 F words. Each study trial began with presentation of the source alone for 1,000ms. The word was then presented next to the source for 2,000ms, and then removed for 1,000ms. The cue was then presented next to the source for 3,000ms. A 1,000ms blank screen preceded the start of the next trial.

The test phase began immediately following the study phase. Participants were told to indicate whether they recognized each presented word, regardless of memory cue. Participants were not given any specific instructions regarding the reinstatement of the study sources, but previously studied items were always tested with their original sources. In the test phase, the test sources were presented in the same locations as at study, and test words were similarly presented to the left or right of the source. Each test source presented 20 R, 20 F, and 40 new words (i.e., 20 of the R words presented from the mostly-remember source and 20 of the F words from the mostly-forget source were not tested); all trials were presented in a random order. Participants were asked to provide a tagging response for each presented item (i.e., to press R if they believed the word had been R cued, F if they believed the word had been F cued, and N if they believed it was a new word). Following completion of the test phase, participants were thanked and debriefed. The entire experiment, including obtaining informed consent and debriefing, took less than 60 minutes.

Results

Recognition

Recognition was first analyzed by combining R and F tags (both indicating an “old” response) into a single hit rate. Hit rates are displayed in Table 2, and were analyzed in a 2 (cue: R vs. F) x 2 (source: mostly-remember vs. mostly-forget) repeated-measures ANOVA. There was a significant main effect of cue, with higher hits to R words (M = .82, SE = .03) than F words (M = .66, SE = .03), F(1,32) = 32.12, MSE = 0.024, p < .001, η_p^2 = .501. The main effect of source approached conventional levels of significance, with numerically higher hits to items from the mostly-remember source (M = .76, SE = .02) than from the mostly-forget source (M = .72, SE = .03), F(1,32) = 4.17, MSE = 0.010, p = .050, η_p^2 = .115. The Cue x Source interaction was not significant, F < 1. Importantly, however, false alarms (see Table 2) were significantly higher to items from the mostly-remember test source than to items from the mostly-forget test source, t(32) = 4.44, p < .001, d = 0.77.

Hits and false alarms were used to compute *d′*; mean values are displayed in Figure 2. The *d′* scores were analyzed in a 2 (cue: R vs. F) x 2 (test source: mostly-remember vs. mostly-forget) repeated-measures ANOVA. There was a significant main effect of cue, with higher discriminability of R words (M = 1.95, SE = 0.14) than F words (M = 1.39, SE = 0.10), F(1,32) = 34.47, MSE = 0.298, p < .001, η_p^2 = .519. The main effect of test source was also significant, with performance actually being worse for the mostly-remember test source (M = 1.50, SE = 0.13) than for the mostly-forget test source (M = 1.84, SE = 0.12), F(1,32) = 8.70, MSE = 0.430, p = .006, η_p^2 = .214. The Cue x Test Source interaction was not significant, F < 1.

Hits and false alarms were also used to compute *c* as a measure of response criterion (see Table 2) and values were analyzed in a similar ANOVA. There was a significant main effect of cue, with more liberal responding to R words (M = -.08, SE = .08) than to F words (M = .20, SE = .09), F(1,32) = 34.47, MSE = 0.074, p < .001, η_p^2 = .519. The main effect of test source was also significant, with more liberal responding to items from the mostly-remember test source (M = -.06, SE = .10) than from the mostly-forget test source (*M* = .19, *SE* = .08), $F(1,32) = 8.70$, *MSE* = 0.097, $p < .001$, $\eta_p^2 = .388$. The Cue x Test Source interaction was not significant, $F < 1$.

Tagging accuracy

The overall proportions of tags assigned in each condition are shown in Table 1. Recognition was also analyzed by examining the accuracy of the tag as R or F, given that an item was recognized (i.e., assigned either an R or F tag); mean tagging accuracy is displayed in Table 2. The proportion correct tags were also analyzed in a 2 (cue: R vs. F) x 2 (test source: mostly-remember vs. mostly-forget) repeated-measures ANOVA. The main effect of cue was not significant, $F < 1$, and neither was the main effect of test source, $F = 1.10$. The Cue x Test Source interaction was significant, $F(1,32) = 14.48$, *MSE* = 0.030, $p = .001$, $\eta_p^2 = .312$. Follow-up comparisons showed that tagging accuracy was greater for R items (*M* = .80, *SE* = .02) than F items (*M* = .66, *SE* = .04) for the mostly-remember test source ($t(32) = 2.72$, $p = .010$, $d = 0.47$). Tags were numerically more accurate for F items (*M* = .80, *SE* = .04) than R items (*M* = .70, *SE* = .04) for the mostly-forget test source, although the difference did not quite reach significance, $t(32) = 1.89$, $p = .068$, $d = 0.33$. Thus, when the sources were present at test, participants were more likely to assign an R tag to items presented with the source that had provided mostly R items during study and more likely to assign an F tag to items presented with the source that had provided mostly F items during study. False alarms were more likely to be tagged as F than as R for both the mostly-remember (*M* = .77, *SE* = .04) and the mostly-forget (*M* = .83, *SE* = .04) test source; the proportions of F tags for false alarms did not differ between test sources ($t(33) = 1.09$, $p = .285$, $d = 0.19$).

Discussion

The goal of the current study was to examine how participants use knowledge about source information to inform their recognition test decisions. Results showed that participants adopted a significantly more liberal response criterion when making recognition decisions about items tested with the mostly-remember source than when making recognition decision about items tested with the mostly-forget source. Reinstating the sources at test allowed for the observation of changes in response criterion, such that although hit rates were higher for items presented with the mostly-remember source at test, so too were false alarms. Thus, item discrimination was actually better for items from the mostly-forget test source, an observation that was not possible to make in Hourihan's (2021) experiments with only one type of new item, and therefore only one false alarm rate that was not associated with either source. The tagging judgments were also consistent with the idea that participants made decisions about the different test source items differently: Participants were more likely to assign R tags to the mostly-remember test source and F tags to the mostly-forget test source, replicating the tagging results observed by Hourihan (2021; Experiment 2). Had participants not been accounting for source information at test, the relative proportions of R and F tags should have been equal for the two test sources. Although the difference was not statistically significant, participants also provided numerically more F tags to false alarms from the mostly-forget test source than from the mostly-remember test source. The pattern in tagging responses also fits with contingency-consistent responding in source memory judgments (e.g., Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012; Spaniol & Bayen, 2002), as participants knew that the two sources had presented differing numbers of memory cues during study, and used this information to assign tags to items presented by those sources at test. Although source knowledge clearly influenced item recognition decisions, it is important to note that DF effects were still observed for both test sources. Intention to remember and the consequent selective rehearsal of some items over others has a powerful effect on item memory.

The results are consistent with a participants using a heuristic strategy to assign tags based on item strength (see Thompson et al., 2011) in conjunction with beliefs about the test source. There are multiple paths to a participant providing an R or F tag. First, one could recollect both the item and the cue from the original study episode, and correctly assign the tag based on this recollection. Second, one could recollect that the item was presented, but not what the cue was. Here, participants could use memory strength and/or the test source information to make a judgment of the most likely cue paired with an item recollected at test. Knowing that sources had presented different numbers of R and F cued words at study, one could reasonably assume that the items tested with the mostly-remember source were more likely to have been R cued than F cued, and items tested with the mostly-forget source were more likely to have been F cued than R cued. Finally, one might not recollect the item or its cue, but may feel a sense of familiarity, and use this level of familiarity in conjunction with the test source information to guess the tag. For the mostly-remember test source, participants know that the source primarily presented R cued items at study that were rehearsed and should feel more strongly familiar, but that some items were F cued, and may lead to a weaker sense of familiarity. This would lead to a bias to provide R tags, based on source knowledge, but also to more liberally assign both R and F tags because participants would expect that some weakly-familiar items were indeed studied, but with F cues. Conversely, for the mostly-forget test source, most items were F cued at study, and, on average, memory for most items from this source should feel weaker. Therefore, the difference in familiarity between an old word that was weakly encoded and a new word would be smaller (than with more strongly encoded items), and participants more carefully evaluated whether they recognized a word, thus using a more conservative criterion.

Although participants also used test source as an indicator of the most likely cue tag, they did also appear to account for memory strength; false alarms, which should correspond to the least familiar items, were more likely to be tagged as F than R, and equivalently so for the two test sources. These results fit well with Dodson and Shimamura (2000), who examined context reinstatement in source memory judgments. They found that reinstating a study context at test improved source memory accuracy, but also increased response bias towards the presented source, regardless of whether it matched or mis-matched the study source. In the current study, participants' item memory judgments were biased towards a response that was consistent with their knowledge of the source study conditions when the source was presented again at test.

Interestingly, McDonough and Gallo (2012) showed that, on old/new recognition tests, participants adopted a more conservative criterion for items that they (falsely) believed would be better learned (large font items). They showed that there were no memory accuracy differences between their two sources when memory was tested using two-alternative forced-choice recognition, which reduces the possible influence of response criteria. In the current experiment, participants essentially completed an old/new test (but with cue-tagging judgments), and a more conservative criterion was used when recognizing items presented by the mostly-forget test source, which was actually more likely to have presented more weakly encoded items during study. That is, despite the two sources presenting an equal number of items during study, participants knew that they were presented with more F words from one source than from the other. Unlike with McDonough and Gallo's (2012) font size manipulation, which influenced memory expectations but not accuracy, selectively rehearsing R cued words does improve memory (and current results showed better recognition of R than F words). Using the heuristic that they had been trying to learn more of the words presented by the mostly-remember source, participants appear to have adopted a more relaxed criterion for calling something old when it was tested with that source. The cue tags they provided were consistent with heuristic decision making based on this source contingency. That is, tagging accuracy showed an interaction between test source and cue, such that participants simply gave more R tags to the mostly-R test source and more F tags to the mostly-F test source. Thus, there is little evidence for actual differences in cue memory accuracy between the two test sources (or between memory cues), but, given the replication of the patterns observed by Hourihan (2021), it seems quite likely that participants primarily made strategic guesses about the memory cue based on their knowledge of the cues the that sources had presented at study.

It is worthwhile noting that McDonough and Gallo (2012) found that participants were more *conservative* when recognizing items they believed were better learned, and here, participants were not only more *liberal* when recognizing items from the mostly-remember test source, but were more liberal when recognizing R than F items in general. Critically, unlike the font size illusion used by McDonough and Gallo, the DF manipulation actually does lead to significant differences in item memory between R and F words, and participants are very much aware of the fact that these cues at study are likely to lead to differences in memory performance at test (Friedman & Castel, 2011). Moreover, participants were not told that they were being tested on an equal number of R and F items from each source, and they therefore likely assumed they were being tested on *all* items from the study phase—twice as many F cued than R cued for the mostly-forget test source. It therefore makes sense to more cautiously evaluate memory for items tested with a source that had mostly presented items one had not been trying to learn, as there would be an expectation for overall memory to be weaker compared to items that had been intentionally rehearsed. Speculatively, informing participants that the test included an equal number of R and F items from each source may reduce or eliminate the differences in response criteria (based on test source, not based on cue) observed in the current experiment.

In summary, the current experiment showed that knowledge about whether a source had provided mostly relevant or mostly irrelevant information changed response criteria at test. When a test source was known to have provided mostly relevant information that was rehearsed, participants were more liberal in item recognition decisions, leading to higher hit rates but also higher false alarm rates. Conversely, decisions were more conservative when the source was known to have provided mostly irrelevant information that was not intentionally learned. Thus, we may be more cautious when evaluating our memories for information associated with a less reliable source; this is likely an adaptive strategy.

References

Bancroft, T.D., Hockley, W.E., & Farquhar, R. (2013). The longer we have to forget the more we remember: The ironic effect of postcue duration in item-based directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 691-699. https://doi.org/10.1037/a0029523

Bayen, U.J., & Kuhlmann, B.G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language, 64*, 1-17. https://doi.org/10.1016/j.jml.2010.09.001

Bayen, U.J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 197-215. https://doi.org/10.1037/0278-7393.22.1.197

Burgess, N., Hockley, W.E., & Hourihan, K.L. (2017*).* The effects of context in item-based directed forgetting: Evidence for “one-shot” context storage. *Memory & Cognition, 45*, 745-754. https://doi.org/10.3758/s13421-017-0692-5

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505. https://doi.org/10.1080/14640748108400805

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dodson, C.S., Schacter, D.L. (2001). “If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review 8*, 155–161. https://doi.org/10.3758/BF03196152

Dodson, C. S., & Shimamura, A. P. (2000). Differential effects of cue dependency on item and source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1023–1044. https://doi.org/10.1037/0278-7393.26.4.1023

E-Prime 2.0. (2015). [Computer software]. Pittsburgh, PA: Psychology Software Tools, Inc.

Fawcett, J. M., & Taylor, T. L. (2008). Forgetting is effortful: Evidence from reaction time probes in an item-method directed forgetting task. *Memory & Cognition, 36*, 1168-1181. https://doi.org/10.3758/MC.36.6.1168

Friedman, M.C., Castel, A.D. (2011). Are we aware of our ability to forget? Metacognitive predictions of directed forgetting. *Memory & Cognition, 39*, 1448. https://doi.org/10.3758/s13421-011-0115-y

Hockley, W. E. (1991). Recognition memory for item and associative information: A comparison of forgetting rates. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock*. (pp. 227–248). Lawrence Erlbaum Associates, Inc.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1321–1330. https://doi.org/10.1037/0278-7393.18.6.1321

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, *36*, 1351–1359. https://doi.org/10.3758/MC.36.7.1351

Hockley, W.E., Ahmad, F.N. & Nicholson, R. (2016). Intentional and incidental encoding of item and associative information in the directed forgetting procedure. *Memory & Cognition, 44*, 220–228. https://doi.org/10.3758/s13421-015-0557-8

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition, 27*, 657–664. https://doi.org/10.3758/BF03211559

Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24*, 202–216. https://doi.org/10.3758/BF03200881

Hockley, W. E., & Niewiadomski, M. W. (2007). Strength-based mirror effects in item and associative recognition: Evidence for within-list criterion changes. *Memory & Cognition, 35*, 679–688. https://doi.org/10.3758/BF03193306

Hourihan, K. L. (2021). The influence of cue probability on item and source judgments in item method directed forgetting. *Memory*, *29*, 1136-1155. https://doi.org/10.1080/09658211.2021.1967400

Hourihan, K. L., & Fraundorf, S.H. (2021). Item method directed forgetting does not affect memory for source. *Manuscript in revision.*

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

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Kuhlmann, B.G., Vaterrodt, B., & Bayen, U.J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1365-1376. https://doi.org/10.1037/a0028147

MacLeod, C.M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1- 57). Mahwah, NJ: Erlbaum.

MacLeod, C.M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. Psychonomic Bulletin & Review, 6, 123-129. https://doi.org/10.3758/BF03210819

Masson, M. E., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203-220. https://doi.org/10.1037/h0087426

McDonough, I. M., & Gallo, D. A. (2012). Illusory expectations can affect retrieval-monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 391–404. https://doi.org/10.1037/a0025548

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. Journal of Experimental Psychology: General, 137(4), 615–625. https://doi.org/10.1037/a0013684

Spaniol, J., & Bayen, U.J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 631-651. https://doi.org/10.1037/0278-7393.28.4.631

Taylor, T.L., & Hamm, J.P. (2018). A grand memory for forgetting: Directed forgetting across contextual changes. Acta Psychologica, 188, 39-54. https://doi.org/10.1016/j.actpsy.2018.05.009

Taylor, T.L., Cutmore, L., & Pries, L. (2018). Item-method directed forgetting: Effects at retrieval? *Acta Psychologica*, *183*, 116-123. https://doi.org/10.1016/j.actpsy.2017.12.004

Thompson, K.M., Fawcett, J.M., & Taylor, T.L. (2011). Tag, you’re it: Tagging as an alternative to yes/no recognition in item method directed forgetting. Acta Psychologica, 138, 171-175. https://doi.org/10.1016/j.actpsy.2011.06.001

Zacks, R. T., Radvansky, G., & Hasher, L. (1996). Directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 143-156. https://doi.org/10.1037/0278-7393.22.1.143

Table 1

Numbers of Words Studied and Tested with Each Source and Mean Proportions of Tagging Responses

	Condition					
	Mostly-Remember Source			Mostly-Forget Source		
	Remember	Forget	New	Remember	Forget	New
Numbers of words						
Study	40	20	-	20	40	-
Test	20	20	40	20	20	40
Proportion Tags						
"R"	**.67**	.24	.09	**.57**	.14	.05
	(.03)	(.03)	(.03)	(.04)	(.03)	(.01)
"F"	.16	**.44**	.19	.23	**.50**	.22
	(.02)	(.03)	(.02)	(.03)	(.03)	(.04)
"N"	.17	.32	**.72**	.20	.36	**.73**
	(.02)	(.03)	(.04)	(.03)	(.04)	(.04)

Note. Standard errors are shown in parentheses below their respective means. Bolded numbers indicate correct tags.

Table 2

Recognition Performance for Remember and Forget Cued Words Based on Test Source

	Condition			
	Mostly-Remember Source		Mostly-Forget Source	
	Remember	Forget	Remember	Forget
Hits	.83	.68	.80	.64
	(.02)	(.03)	(.03)	(.04)
False Alarms	.28		.16	
	(.04)		(.02)	
Response Bias (*c*)	-.20	.08	.05	.33
	(.09)	(.11)	(.08)	(.09)
Tagging Accuracy [a]	.80	.66	.70	.80
	(.02)	(.04)	(.04)	(.04)

Note. Standard errors are shown in parentheses below their respective means.

[a] Computed as the proportion of items called either "R" or "F" that were assigned the correct tag.

Figure 1

Example Study and Test Trials

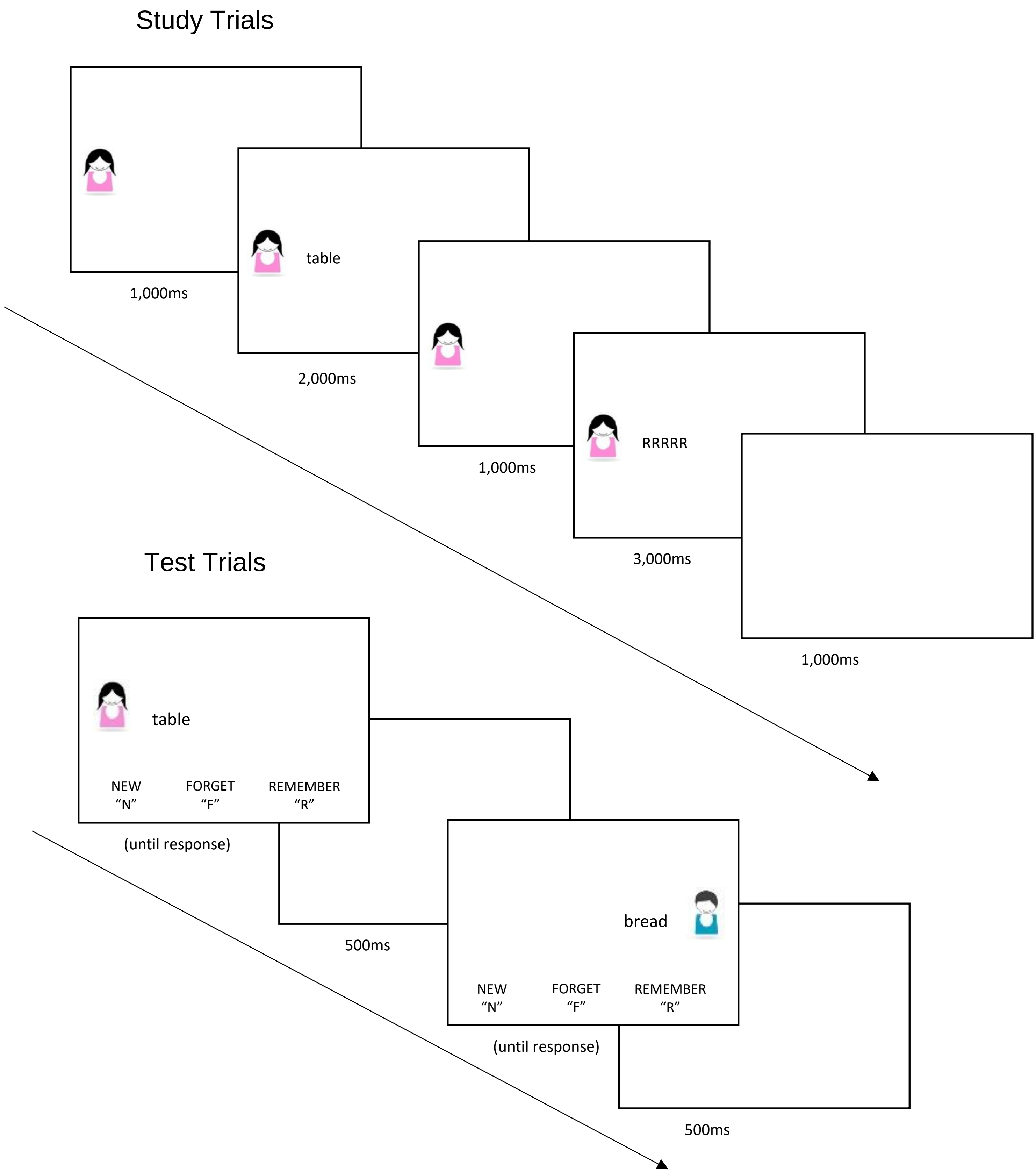
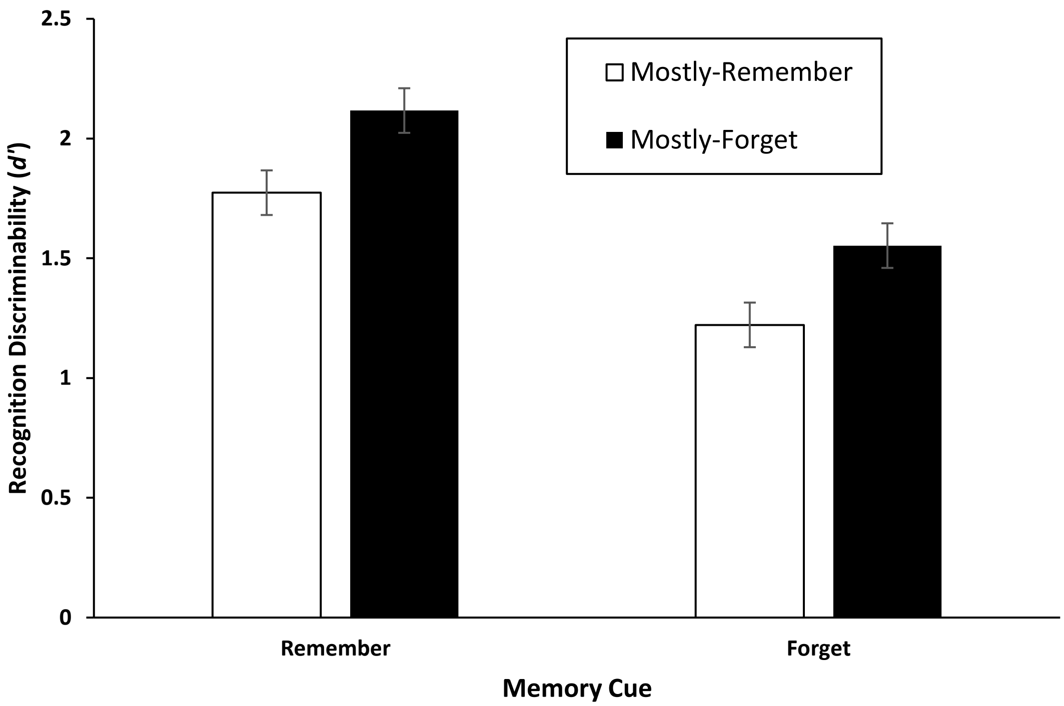


Figure 2

Mean Recognition Discriminability (d') as a Function of Memory Cue and Source



Note. Error bars represent 95% confidence intervals of the means (Masson & Loftus, 2003)