The Reality of Directed Forgetting in the Item-Method Paradigm:

Suppression, not Selective Search or Decay.

by

Whitney Anne Hansen

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2011 by the
Graduate Supervisory Committee:

Stephen Goldinger, Chair
Tamiko Azuma
Gene Brewer
Donald Homa

ARIZONA STATE UNIVERSITY

May 2011

ABSTRACT

It has been suggested that directed forgetting (DF) in the item-method paradigm results from selective rehearsal of R items and passive decay of F items. However, recent evidence suggested that the passive decay explanation is insufficient. The current experiments examined two theories of DF that assume an active forgetting process: (1) attentional inhibition and (2) tagging and selective search (TSS). Across three experiments, the central tenets of these theories were evaluated. Experiment 1 included encoding manipulations in an attempt to distinguish between these competing theories, but the results were inconclusive. Experiments 2 and 3 examined the theories separately. The results from Experiment 2 supported a representation suppression account of attentional inhibition, while the evidence from Experiment 3 suggested that TSS was not a viable mechanism for DF. Overall, the results provide additional evidence that forgetting is due to an active process, and suggest this process may act to suppress the representations of F items.

To my parents, Pete and Thea, who have provided unconditional support throughout my life and this process. Thank you.

ACKNOWLEDGMENTS

I thank the members of my committee, Tamiko Azuma, Gene Brewer, Donald Homa and Stephen Goldinger for feedback on theory and methodology. I also thank Stephen Goldinger for his guidance, encouragement, and mentorship on this project and many others. Finally, I thank my colleagues, Anthony Barnhart, Michael Hout and Megan Papesh for their help with stimuli and programming, as well as their continued support and friendship. This dissertation was funded in part by a dissertation fellowship awarded from the Graduate College of Arizona State University.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure	Page

The reality of directed forgetting in the item-method paradigm:
Suppression, not selective search or decay.

Often forgetting is described as a failure of an imperfect memory system, but it can also be quite useful. Forgetting allows for information that is distracting or misleading to be discarded, consequently freeing up cognitive resources for goal-relevant tasks. This process has been studied across psychology: in the clinical domain through work on repression, in social psychology through judgment research, and in cognitive psychology, most notably, through a directed forgetting (DF) paradigm. In DF experiments, participants must update and control the contents of memory, remembering certain items and forgetting others (Macleod, 1998). To illustrate the utility of forgetting, consider one early experiment. Muther (1965) presented participants with letters to remember across three conditions. In the critical condition, participants were presented 12 letters that were to be remembered (TBR), and 12 letters to be forgotten (TBF). The other conditions acted as controls: in the first control condition only the 12 TBR letters were presented, in the second control condition all 24 letters were presented as TBR. Muther found that memory for the critical 12 TBR items was worst in the condition with 24 TBR items, better in the condition with 12 TBR and 12 TBF items, and best in the condition where only the 12 items were presented. Muther concluded that allowing participants to forget some letters improved memory for the remaining letters, but forgetting came at a cost, in that recall was not as good as it was when the TBF items were omitted. This provided early evidence that intentional forgetting of unwanted information can occur, but is an imperfect process.

Muther's experiment provided one of the first systematic examinations of forgetting, and paved the way for additional studies. Since his seminal work, a variety of testing procedures have been used, but two prominent methods have emerged. Across both designs, participants are presented stimuli (most often words) and told to remember (R) half of them for a later memory test. They are told they can forget (F) the other half because they will not be tested. After all the stimuli are presented, a recognition or recall test is administered and participants are to respond "old" to both R and F words, despite the initial instructions. The two methods differ in their administration of the R and F cues (Basden, Basden & Gargano, 1993). Under the *item-method* design, participants are shown a series of stimuli each followed by a cue signaling the item should be remembered or forgotten. In the *list-method*, participants are stopped half way through a list of items and told that the previous items were "practice" and they should now be forgotten. The second half of the list is then presented, and participants are told to remember those items. The *directed forgetting effect* refers to the finding that R items are correctly recognized (or recalled) more often than F items, suggesting that the memory cue was effective in inducing forgetting. The (R – F) difference provides a measure of the size of the effect, and can be altered by changing testing conditions.

The DF effect is taken as evidence that forgetting can and does occur, but the underlying mechanism remains unclear. Several theories have been proposed, and will be discussed next. Before presenting the theories, it should be noted that the DF effect has been replicated in a variety of testing situations, and there is evidence that it is not the result of demand characteristics to underreport remembered F items. Compelling behavioral evidence comes from Kausler and Settle (1975) who found that homophone errors are greater to R words, relative to F words, underscoring the notion that F words are less accessible in memory and arguing against demand characteristics. MacLeod (1999) tested the role of demand characteristics directly. He examined DF of words across item- and list-method designs, using recognition, recall, and source monitoring testing methods. First, participants completed an initial recall task, where they were instructed to recall all R and F items that they remembered. After the initial recall test, a monetary incentive was provided for any additional F words that could be recalled. Next, an “old”/”new” recognition test was presented for targets and foils, followed by a final source-monitoring task for the studied words, wherein participants were asked to indicate which instruction (R or F) was associated with each word at encoding. The typical DF effect was observed across both the item and list methods and across all testing procedures. Importantly, when given the incentive to recall additional F items, participants were unable to do so, recalling, on average, less than one additional F word. His results suggest that participants are not consciously failing to report F words even though they are remembered, undermining explanations that include demand characteristics or reduced motivation to retrieve F items.

Theories of Directed Forgetting

Active Erasure

The first theory of DF, active erasure (Bjork, LaBerge, & Legrand, 1968) states that items are plucked from short-term memory (STM), and removed completely. This theory was short lived, because there is ample evidence that at least some F items make their way into long term memory (LTM). For example, F letters are more likely to appear in a recall test than other letters that were not presented (Muther, 1965), participants often call F words “old” when instructed to recognize or recall them (Bjork, 1970), and F items often show priming effects on implicit tests (Basden, et al.1993). Thus, the active erasure hypothesis was soon discarded, and other theories emerged.

Selective Rehearsal

The selective rehearsal theory states that items are held in working memory until the R/F cue appears. At the cue, a selective rehearsal process is engaged for R items, but not for F items, which eventually decay from memory. Although many have espoused this perspective (Basden, et al.1993; Bjork, 1972), much of the supportive evidence comes from work done by Wetzel and colleagues. Using an item-method design, Wetzel and Hunt (1977) manipulated the amount of time that followed the R/F cue, thus manipulating the time participants had to engage in an elaborative rehearsal processes. They found that increasing this post-cue interval lead to better memory performance for R items, but did not affect memory performance for F items. In a second experiment they prevented post-cue rehearsal and found that this distraction affected performance for R words, but not for F words. In another study, Wetzel (1975) manipulated the time between the word and the R/F cue, the period during which it is assumed that words are simply held in WM. He found that increasing this interval actually lead to poorer memory for R words. Thus, the early evidence was consistent with the theory that R words are selectively rehearsed, while F items are ignored and left to decay from memory.

Inhibition

Several theories espouse inhibitory processing as a key component of intentional forgetting, but these theories advocate various roles for inhibition. One theory, *retrieval inhibition*, suggests that an inhibitory process prevents the retrieval of F items at testing. Compelling evidence for this theory comes from Geiselman and Bagheri (1985). Using an item-method design, they presented words that were either TBR or TBF, and the classic DF effect was observed on an initial recall test. They then presented most of the words again, but this time all were TBR. During a second recall test, memory improved much more for the F items, relative to the R items, though the DF effect still appeared. They suggested that F words benefited from a release of retrieval inhibition that was provided via the second encoding session, in addition to the added rehearsal benefits that both R and F words received. A series of control experiments were included to eliminate alternative explanations including variations in item difficulty and preferential rehearsal of F items during the second recall test. One additional explanation stated that the advantage for F items on the second recall test appeared because more F items were available from the first test. To test this hypothesis, they presented only 10 unrecalled R items and 10 unrecalled F items during the second encoding phase, so that opportunities for recall were equal across R and F items. They found that significantly more F items were recalled during the second testing session, relative to R items. They conclude, “F items suffer from retrieval inhibition in addition to insufficient encoding…It is hypothesized that a command to forget initiates a process that serves to inhibit or block access routes to the target items in episodic memory” (p 62).

A second hypothesis, *attentional inhibition* (Zacks, Radvansky, & Hasher, 1996), states that inhibition at encoding can be used to suppress the processing of goal-irrelevant information. In previous work, Hasher and Zacks (1988) showed that older adults have a reduced ability to inhibit goal-irrelevant activity. If attentional inhibition is important in DF, older adults should have a reduced ability to forget words they were instructed to forget. To examine this, Zacks et al. (1996) selected words belonging to specific categories (animals) as stimuli. Within a category, some words were TBR and some TBF. They found that older adults recalled fewer items overall, recalled fewer TBR items, and were more likely to incorrectly report TBF items. Furthermore, older adults had a more difficult time differentiating between TBR and TBF words belonging to the same category. In a second experiment, they had older and younger adults perform a Sternberg task with DF instructions (Sternberg, 1966), and found that older adults showed a larger slow-down in response times (RTs) to TBF words. They concluded that the reduced ability of older adults to inhibit rehearsal of F words led to slower RTs in the Sternberg task, as well as increased F intrusions in the traditional DF test.

Recently, Fawcett and Taylor (2008) compared the selective rehearsal and attentional inhibition theories. They added a probe-detection task to an otherwise typical item-method DF paradigm to measure attentional demands across the post-cue delay. During the probe task, participants were interrupted from their primary responsibility (in this case, remembering or forgetting), and asked to respond to the probe as quickly as possible. RTs to the probe reflect the cognitive effort required by the primary task. Therefore, RTs on the probe task should be slower when processing in the primary task is more effortful. The selective rehearsal theory assumes that F representations are allowed to simply decay overtime, leading to their poor recollection (Basden, et al.1993). Thus, under this hypothesis, RTs to the probe should be slower for R words, which are rehearsed, relative to F words. Conversely, the attentional inhibition theory (Hasher & Zacks, 1988; Zacks et al., 1996) suggests that an effortful process acts to suppress F items, predicting that probe RTs to F words should be similar to, or even slower than, R words.

Consistent with the item-method design, Fawcett and Taylor (2008) presented participants words followed by memory cues. At one of three stimulus onset asynchronies (SOAs; 1400 ms, 1600 ms, 2600 ms) after the memory cue appeared, a probe was presented and participants pressed the space key as soon as it was detected. At the end of the encoding trials, a “yes/no” recognition task was used to ensure performance on the primary DF task was maintained. Although the classic DF effect was observed, RTs for probe detection were the dependent variable of main interest. They found that probe RTs were significantly longer following F instructions at 1400 and 1600 ms SOAs, but not at the 2600 ms SOA. This not only suggests that forgetting is effortful, but suggests that at least initially, forgetting is *more* effortful than remembering. This pattern is inconsistent with a selective rehearsal account, but is consistent with the attentional inhibition hypothesis.

Tagging and Selective Search

Several theories have suggested that DF is the result of a selective search process at retrieval, including *selective search* (Epstein, Massaro & Wilder, 1972; Shebilske, Wilder, & Epstein, 1971), *segregation* (Bjork, LaBerge, & Legrand, 1968), and *set differentiation* (Bjork, 1970; Elmes, Adams & Roediger, 1970; Horton & Petruk, 1980). Although the names differ, the imperative mechanisms are similar. First, R and F items are tagged and organized at encoding. At retrieval, a selective search process is engaged through one set of stimuli (most often R). Selective searching facilitates retrieval of the target set, while reducing interference from the unwanted set. Exactly how this is achieved differs slightly across the theories. According to Bjork (1970), participants use F cues to segregate F and R items, and then to rehearse the R items. These processes are codependent in that the segregation requires differential processing, and the differential processing further encourages segregation. Epstein and Wilder (1972) acknowledged that selective rehearsal may contribute, but if it is operating, it serves the purpose of keeping R words active for search. Elmes et al. (1970)

hypothesized that F items are tagged at encoding, and then actively suppressed. At retrieval, a search is engaged for R items only, and F items are avoided. Across all theories, tagging leads to differential encoding of R and F items, which allows F items to be avoided at testing. Thus, several of these selective search theories combine aspects of the previous encoding theories such as selective rehearsal or inhibition with a targeted search process at testing. I will use the term *tagging and selective search* (TSS) because I feel it most accurately represents the crucial mechanisms: tagging and organizing of R and F items at encoding, and selective search of items at retrieval.

Early evidence supporting selective search came from experiments examining DF in STM. Epstein, Massaro and Wilder (1972) presented participants with three consonant-vowel-consonant (CVC) – word pairs. Following a short break, a second set of three CVC-word pairs was presented, followed by a memory cue and a post-cue rehearsal period. At testing, participants were shown the CVC stimulus and asked to retrieve the corresponding word. Instead of R or F cues, the memory cue was FIRST, SECOND or EITHER, and reflected the set of CVC-word pairs from which the target would appear. If FIRST or SECOND appeared, participants could focus rehearsal *only* on the appropriate set (ONLY trials), but if an EITHER cue appeared, all six CVC-word pairs had to be maintained. In other words, the ONLY trials provide an opportunity for participants to forget half of the studied items. Because forgetting some items facilitates memory for TBR items (Muther, 1965), memory performance was expected to be better on ONLY trials, relative to EITHER trials (the *only effect*; Shebilske, Wilder, & Epstein, 1971).

According to selective search, memory for R items is better than memory for F items because the search set is restricted to R items. Thus, the size and composition of the search set should have an effect on memory performance. Epstein et al. (1972) manipulated search set by including two testing conditions. In the matching condition, participants were presented three words, and asked to select the correct word. This equated the size of the search set across ONLY and EITHER trials, as both contained three items. However, during the recall condition, the set size was larger for EITHER trials (six items), relative to ONLY trials (three items). If selective search was correct, a Cue x Testing Method interaction was predicted to emerge: The *only effect* should emerge during recall but not during matching conditions. As predicted, they found that performance was better during recall on ONLY trials, relative to EITHER trials, but no difference appeared in matching. To rule out the possibility that differential processing at retrieval drove this interaction, they ran a second experiment during which all six words were presented in the matching condition, and the same interaction emerged. Epstein et al. concluded that the patterns were best explained with a selective search account.

Using a similar methodology, Homa and Spieker (1974) evaluated the selective search account by examining RTs at testing. They presented CVC-word pairs in sets, which were separated by an unfilled interval. The size of the sets varied (2, 4, or 6) across trials, and remembering and forgetting were encouraged via FIRST, SECOND, or EITHER cues. Selective search predicts that RTs should be shorter, overall, when only half of the list needs to be searched (ONLY trials). Furthermore, RTs should increase as the TBR set size increases. Finally, if the search process mirrors a serial-exhaustive search (Sternberg, 1966), then the slope increase for ONLY trials should be half of the size of the slope for EITHER trials. Consistent with the first two predictions, RTs were faster for ONLY, relative to EITHER trials, and RTs increased as the set size increased. The slope for ONLY trials was not half the slope for EITHER trials, leading Homa and Spieker to conclude that 1) selective search is operating, but 2) the search process is likely to be serial, self-terminating, and not serial-exhaustive.

TSS is an ironic theory of forgetting, in that it implies that F items are not actually forgotten: F items are stored, but ignored at encoding and avoided at retrieval. Although it may be counterintuitive, the idea that R and F items are stored and organized is consistent with several important findings, and these results are difficult to reconcile with other theories of DF. For example, F items interfere with each other but not with R items, as selective search predicts (Reitman, Malin, Bjork, & Higman, 1973). Recall of F items improves if the participants know it is an F item (Epstein & Wilder, 1972). Instructions to organize F items facilitates retention of R items, presumably because is facilitates segregation of the F and R items (Geiselman, 1975). Finally, TSS is the only theory that explicitly states that F items are retained yet underreported. This provides a potential explanation for results that argue against true forgetting, such as priming effects for F items (Basden et al., 1993).

Multiple Mechanisms for Directed Forgetting

Aside from *active erasure*, the previous theories all provide viable explanations for directed forgetting. Each theory has garnered evidentiary support, although no one theory has been able to account for the wide variety of (sometimes conflicting) data that have emerged. In an attempt to make sense of the variety of data and clear up inconsistencies across experiments, several previous experiments have directly compared the competing theories. For example, MacLeod (1989) examined DF effects in explicit and implicit tests in an item method design. The selective rehearsal and retrieval inhibition accounts of DF make different predictions about when DF effects should occur. Selective rehearsal states that R items receive rehearsal benefits, relative to F items. MacLeod noted that if this were true, DF effects should not be observed in implicit tests, as previous evidence suggested that implicit tests were not influenced by elaborative encoding (Jacoby & Dallas, 1981). On the other hand, if retrieval inhibition were correct, then DF effects could be observed on both implicit and explicit tests, as the retrieval mechanisms are similar across the two testing methods. MacLeod found DF effects in both explicit (recognition, free recall) and implicit (fragment completion, lexical decision) tasks across two experiments, consistent with the retrieval inhibition account.

Basden et al. (1993) questioned MacLeod's results and his logic, and in the process provided a critical insight into DF. Specifically, they suggested that the various methods and testing measures used in DF may invoke different remembering and forgetting mechanisms. They pointed out that the item- and list-methods had been used interchangeably, but the results were often inconsistent. For example, DF effects were observed in recognition tests when the item-method was used, but were not always observed with the list-method. However, DF effects were consistently found for both item- and list-method when a recall test was administered. A second discrepancy arose when comparing performance on implicit memory tests. DF effects were observed when using the item method for some implicit tasks, including lexical decision and word fragment completion (MacLeod, 1989), but were not observed in others, including a word association test, and fragment completion (Basden, et al.1993).

These discrepancies led Basden et al. (1993) to hypothesize that the mechanisms responsible for DF across the item- and list-methods and explicit and implicit tasks were different. They argued that selective rehearsal is plausible in the item-method, but makes less sense as an explanation of the list-method because the R/F cue appears sometime after rehearsal had occurred. On the other hand, the list-method seems better suited to a retrieval inhibition account, wherein a subset of stimuli can be grouped together and then suppressed. If their reasoning was correct, it may explain the conflicting findings across recognition and recall tests. For example, as noted previously, DF effects are often not observed when recognition tests are used under list-method encoding. If retrieval inhibition is the driving force in the list method, re-presenting the items (as is done in recognition tests) should lift the retrieval inhibition, leading F words to be recognized almost as well as R items. Recall, on the other hand, requires a strategic search of memory. Performance would benefit from selective rehearsal of R items as well as retrieval inhibition of F items. Thus, DF effects should be robust for recall under both item- and list-methods. They tested their hypothesis across four experiments, using list- and item-method at encoding and recall and recognition at testing. The explicit memory test results were consistent with their predictions: DF effects appeared under the item- and list-methods when recall tests were given, but only under the item-method when recognition tests were given.

They also compared implicit and explicit testing procedures, including fragment completion and fragment cued recall tests. They disagreed with MacLeod's assumption that retrieval processing was equivalent across implicit and explicit tasks. Arguing from a Transfer Appropriate Processing perspective (TAP; Roediger, 1990), they suggested that to the extent that processing at encoding and testing match, memory performance will improve. Thus, if a conceptually-driven encoding task is presented (generating words, elaborative rehearsal), performance on a conceptually-driven test (recall) will improve. Importantly, if one adopts a TAP framework, retrieval inhibition and selective rehearsal theories make different predictions about DF effects in implicit tasks. According to retrieval inhibition, equivalent performance should be observed on both conceptually-driven (fragment cued recall) and data-driven (stem completion) tests. Conversely, selective rehearsal suggests that the DF effect should emerge on conceptually-driven tests (with elaborately encoded R items remembered better than F items), but not on data-driven tests. They failed to find the classic DF effect for implicit tests when the list method was used, but the effect did appear under the item-method. This is consistent with the idea that selective rehearsal is contributing to performance in the item-method, but not responsible for DF in the list-method. They argue MacLeod's tests were data-driven tests, leading to his observed DF effects. Overall, their results suggested that selective rehearsal and retrieval inhibition both contribute to DF, but in different ways based on the encoding and testing methodologies that are used.

The Current Experiments

Basden et al.'s (1993) results were as surprising as they were compelling. Across four experiments, their evidence suggested that selective rehearsal was responsible for DF in the item method, while retrieval inhibition was responsible for the list-method. Other theories of DF, such as attentional inhibition and TSS, have been largely ignored in the item-method literature as a result of their work, despite a variety of evidence supporting each theory's claims. The current experiments re-examined the contributions of TSS and attentional inhibition in DF in an item-method design. Across three experiments, various procedures, stimuli, and manipulations were introduced to examine the basic tenets of these two theories. Experiment 1 compared competing predictions made by TSS and attentional inhibition theories. Experiment 2 focused on the attentional inhibition theory and evaluated two explanations for inhibition. Experiment 3 scrutinized the TSS theory by comparing performance in DF conditions to performance in conditions during which a TSS approach was encouraged. The primary goal of these experiments was to examine the role that inhibition and TSS might play in DF. However, some of the current results speak to the selective rehearsal theory of DF, and when appropriate this theory will be discussed as well.

Experiment 1

TSS and attentional inhibition are two significant theories of DF, but thus far direct comparisons of the two are scarce. Experiment 1 provides such a direct test by including manipulations that can differentiate between the competing theories so that strong inferences about the correct mechanisms can be made (Platt, 1964). The theories make different predictions about the effect that encoding manipulations should have on the size of the DF effect. According to TSS, manipulations that improve encoding should facilitate tagging and organizing of R and F items. This should result in poorer memory for F items. In other words, the discrepancy between memory for R and F items, the *DF effect*, should be larger when tagging is facilitated via encoding benefits. In contrast, attentional inhibition states that manipulations that are known to improve encoding should improve memory for R items, but make it more difficult to suppress the representation of F items. This should improve memory for F items at testing, resulting in a smaller DF effect. These predictions were pitted against one another in the current design, through two manipulations: Level of Processing (Craik & Lockhart, 1972) and timing of the memory cue.

Manipulating the level of processing (LOP) that occurs at encoding has consistent and robust effects on later retention. When asked to orient to different aspects of a stimulus, memory is better for orienting tasks that require elaborate processing. For example, subsequent memory performance will be better if one was asked to make a judgment about the meaning of a printed word at encoding, relative to the sound of the word, or the characteristics of the font. The classic finding is that deeper, semantic processing leads to a more durable memory trace and results in better memory performance over shallow, phonological or perceptual processing. This result has been replicated numerous times across a variety of encoding and testing instructions (Craik & Tulving, 1975).

LOP has been included in DF tests previously, but the evidence is mixed. Wetzel (1975) first included LOP in an item-method design. He presented four lists of words aurally, and each word within a list was followed by a tone (signaling that item was TBR), or a click (TBF). At encoding, participants were asked write down the target word (rehearsal condition), write down a word that rhymed with the target word (perceptual condition), or write down an adjective that modified the word (semantic condition). Four immediate free recall (IFR) tests followed the presentation of each list, along with a final free recall (FFR) test and a recognition memory test. Across all three memory tests, the typical DF effect was observed, with R items recalled more often than F items. No main effect of LOP was observed, but the LOP x Memory Cue interaction was significant in both IFR and FFR. Surprisingly, Wetzel found that recall for R words was significantly better in the rehearsal and rhyming conditions, relative to the semantic condition, a result opposite of the typical LOP effect. Additionally, LOP had no effect on F items. In Experiment 2, Wetzel varied the time between the target word and the memory cue (1 second), relative to Experiment 1 (5 seconds), leaving all other manipulations intact. He found that the closer temporal cue enhanced memory for R items, but not F items. More importantly for the current purposes, the LOP and Memory Cue results were identical to

Experiment 1. Thus, across two experiments Wetzel failed to find a main effect of LOP, and the LOP x Memory Cue interaction that emerged was opposite of the predicted trends and previous evidence (Craik & Lockhart, 1972; Johnston & Jenkins, 1971).

The unusual pattern of results obtained by Wetzel (1975) led Horton and Petruk (1980) to revisit the effect LOP has on DF. Several factors distinguish their design from Wetzel's (1975). First, their encoding tasks were slightly modified: Participants were asked to generate a word that began with the same letter as the target word (structural), a word that rhymed with the target (phonemic), or a word that belonged to the same conceptual category (semantic). Second, the structure of the word lists was manipulated so that words within a list were unrelated, related so that all items of a given category were either TBR or TBF, or related so that category members were divided equally among R and F cues. Thus, in the two related conditions, the memory cue could be *orthogonal* to the semantic category, or *non-orthogonal*, wherein the memory cue and the semantic category were correlated. Finally, they varied when the cue was presented, relative to the orienting decision, to determine if the temporal proximity of the cue would predict later memory performance. The cue was always presented after the word, but could be presented either before (pre-cue) or after (post-cue) the LOP task.

Horton and Petruk presented three lists of words to participants. After each list, an IFR task was given, and ultimately a FFR test and a modified recognition test were included. During recognition, participants were given a subset of the previously presented items and asked to add the type of encoding task and memory cue that accompanied each word. Given the variety of testing measures and manipulated independent variables they included, their results were extensive, but only those results most relevant to the current discussion are included. Horton and Petruk observed the classic DF effect in IFR, FFR, and recognition tests, with R items better remembered than F items. Unlike Wetzel (1975), they found the typical LOP effect across all three tests: words processed under semantic encoding conditions were better remembered, relative to phonemic and structural encoding conditions. Furthermore, this pattern was observed for both R and F items. The Memory Cue x LOP interaction was significant at IFR and recognition but not FFR, and indicated that the size of the DF effect was larger for deep, relative to shallow, encoding conditions. Finally, the Cue Timing x Memory Cue interaction was significant at IFR but not FFR. Pre-cued items were recalled better than post-cued items, if the memory cue was R, but not when the memory cue was F.

To summarize, the extant literature on LOP effects in DF is contradictory: Wetzel failed to find the typical LOP effect, while Horton and Petruk (1980) observed the effect for both R and F items. Thus, the effect that LOP has on DF is unclear. Additional tests of this manipulation are necessary to determine if and how depth of encoding influences remembering and forgetting.

Overview of the Current Experiment

The current experiment provides an additional examination of LOP effects on DF, while including several notable modifications. Unlike previous studies where participants were asked to generate words related to the target, participants in the current design made a judgment about each word (Craik & Tulving, 1975). In the deep encoding condition, participants were asked if the word described an object that was animate or inanimate. In the shallow condition, they were asked if the word contained the letter E. In addition, a second manipulation, Cue Timing, was included. Traditionally, the R/F cue is presented after the word, but in the current experiment it was presented before the word on half of the trials. As mentioned, Horton and Petruk (1980) included a timing manipulation in their study, wherein the cue was presented before or after the LOP decision was made. Importantly, in their experiment, the cue was *always* presented after the word. In the current experiment, the cue was presented either before the word or after the orienting decision. This change from Horton and Petruk's method is not trivial. In their experiment both R and F words were processed initially as participants awaited the memory cue and orienting task. Whatever processing occurred to differentiate R and F words had to overcome this initial processing. In the current experiment, differential processing of R and F words could begin immediately (with one caveat discussed below). Finally, a recognition memory test was used instead of free recall. Recognition is less dependent on retrieval mechanisms, which should enhance any effects that occur at encoding, including cue timing and LOP (Wetzel, 1975). Although both Wetzel and Horton and Petruk included recognition test eventually, their tests occurred after immediate and final recall tests, which may have contaminated their recognition data.

As mentioned, the attentional inhibition and TSS theories make different predictions about how the current manipulations should affect subsequent remembering and forgetting, and thus the size of the DF effect (see Table 1). First, the predictions for LOP will be discussed. According to attentional inhibition, deeper processing should make it more difficult to suppress items that are paired with F cues, leading to better recognition of F items at testing. This result would lead to a smaller DF effect under deep encoding conditions, relative to shallow conditions. However, deeper encoding may improve memory for R items, as well. If the relative improvement for R and F items is similar, then the size of the DF effect would be equal across the two levels of LOP, but if the size of the benefit for R items is smaller than the size of the suppression cost to F items, then the DF effect will be smaller. Thus, attentional inhibition states that the DF effect under deep encoding should be equal to or smaller than the DF effect under shallow encoding. According to TSS, deeper processing should result in enhanced memory for R items and poorer recognition of F items, relative to shallow encoding. This result may seem counterintuitive, but according to TSS, deep processing should reduce memory for F items, as enhanced tagging would result in F items being ignored at encoding and avoided at testing. Thus, the size of the DF effect is expected to be larger under deep processing, relative to shallow processing conditions.

With respect to Cue Timing, the predictions are not orthogonal, but opportunities for discriminating between the theories do exist. According to attentional inhibition, an F cue presented before the word should engage the inhibitory process immediately, so that the absolute minimal amount of processing occurs. Although ideal, this is not possible in the current experiment, as the orienting task requires some attention to be dedicated to every word. Therefore, inhibitory processing of F items should be consistent across pre- and post-cue trials, translating to equivalent memory performance. However, a pre-cue should improve memory for R words, as memorization can begin immediately once the word appears. Hence, the DF effect should be larger for pre-cue, relative to post-cue conditions. According to TSS, a pre-cue would be beneficial for R items, and detrimental for F items, as the pre-cue allows both types of words to be immediately tagged and sorted. Efficient organization should lead memory for R items to be improved, and F items to be diminished (Horton & Petruk, 1980). Therefore, TSS also predicts that the pre-cue should result in a larger DF effect, relative to the post-cue. Although both theories predict that the size of the DF effect should be larger for pre-cues, relative to post-cues, they make different predictions about how F items will be affected. According to TSS, hits to F items should be lower with pre-cues, relative to F items for post-cues, while attentional inhibition predicts no change.

These predictions were testing in a fully counter-balanced 2 (LOP) x 2 (Cue Timing) x 2 (Memory Cue) design. Four encoding conditions were included (Deep before, Deep after, Shallow Before, Shallow After), and memory cue (R, F) was manipulated within each condition. One recognition test was given, during which participants responded “old” to both R and F words. Unlike previous examinations of LOP, this design was completely within subject, providing the strongest opportunity for any relevant effects to appear.

Table 1

Summary of the predictions made by the Tagging and Selective Search (TSS) and Attentional Inhibition theories for R, F and DF effects as a function of the level of processing, cue timing, and memory cues.

		R items	F items	DF effect
LOP	TSS	$R_{deep} > R_{shallow}$	$F_{deep} < F_{shallow}$	$(R-F)_{deep} > (R-F)_{shallow}$
	Inhibition	$R_{deep} > R_{shallow}$	$F_{deep} > F_{shallow}$	$(R-F)_{deep} < (R-F)_{shallow}$ $(R-F)_{deep} = (R-F)_{shallow}$
Cue Timing	TSS	$R_{pre\text{-}cue} > R_{post\text{-}cue}$	$F_{pre\text{-}cue} < F_{post\text{-}cue}$	$(R-F)_{pre\text{-}cue} > (R-F)_{post\text{-}cue}$
	Inhibition	$R_{pre\text{-}cue} > R_{post\text{-}cue}$	$F_{pre\text{-}cue} = F_{post\text{-}cue}$	$(R-F)_{pre\text{-}cue} > (R-F)_{post\text{-}cue}$

Method

Participants

Ninety-nine students enrolled in an introductory psychology course at Arizona State University participated for partial course credit. Data from five participants were excluded from analysis, because accuracy on the memory task fell below chance (50%), leaving data from 94 participants for final analyses.

Materials

A total of 268 words were included, of which 12 were dedicated to practice trials. The remaining 256 were selected from the MRC database (Wilson, 1988) and various internet websites (http://en.wikipedia.org/wiki/List_of_animal_names), so that each word fit into one of the four categories required for the LOP manipulation: Animate with an E, Animate without an E, Inanimate with an E, Inanimate without an E (see Appendix A). Because it was important that the semantic and orthographic characteristics of the words be equally represented across the various encoding conditions, word lists were pseudo-randomly created prior to the experiment according to the following procedure. First, half of the words were assigned to be targets, and half lures. From the target list, four lists of 32 words were then created for the four encoding conditions (deep before, deep after, shallow before, shallow after). Each list contained eight words that belonged to each of the four word categories. To ensure that any obtained results were not due to the pseudo-random creation of the word lists, four versions of the experiment were created, so that each item appeared equally across the encoding conditions. Finally, after the four versions of the experiment were created, half of the targets in each condition were assigned to with an "RRR" cue and half with the "FFF" cue. The assignment of memory cue to word was counterbalanced across conditions, so that each word was associated with a "RRR" cue and "FFF" cue equally across the four versions of the experiment. Stimuli and memory cues were presented centered in black Courier New 18 point font on a gray background.

Procedure

The experiment was run in a large testing room containing up to twelve testing stations. The experiment was designed using E-prime software (Psychological Software Tools; Schneider, Eschman, & Zuccolotto, 2002) and presented on 17-in. (16-in. viewable) CRT computer screen. The design was entirely within-subjects, and testing took no more than one hour.

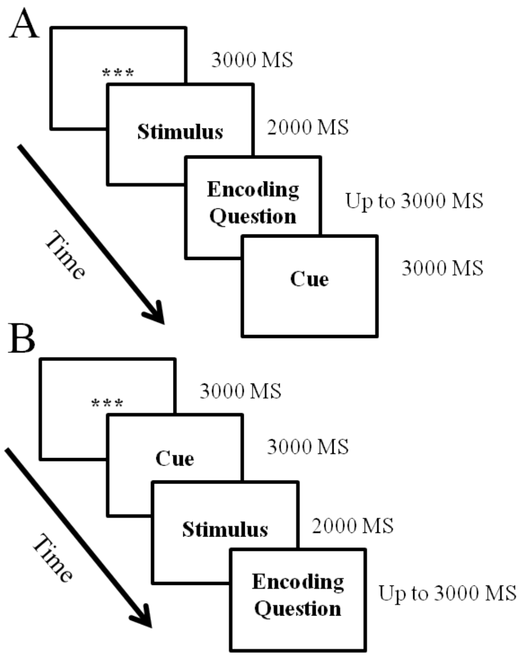

Figure 1. The timing of an individual encoding trial in Experiment 1 when the cue was presented after the word (A) or before the word (B).

As seen in Figure 1, encoding trials included a fixation stimulus (***) that appeared for 3000 ms, a target word for 2000 ms, a decision screen containing the orienting task that appeared until a decision was made or until 3000 ms had elapsed[1], and the memory cue (RRR or FFF) for 3000 ms. In the deep encoding condition, the orienting question was: “Is the word an animate object?”, whereas in the shallow condition it was: “Does the word contain an E?” Across both conditions, the appropriate keystrokes (Yes = V, No = N) were also presented underneath the orienting question. The decision screen always appeared immediately after the target word. The presentation of the memory cue varied according to the Cue Timing condition, appearing either before the word or after

[1] Evidence obtained from pilot tests revealed that reaction times to the decision cue did not differ across groups, thus the timing of the decision cue was allowed to vary. Post-hoc comparisons of the current data confirmed that no RT difference existed across conditions.

the decision screen. The total duration per trial was 8000 ms, plus the time needed to respond to the decision prompt.

Participants began with two sets of practice trials. The first set of four practice trials introduced the LOP manipulation. The second set of practice trials contained both the LOP and memory cue manipulation. Participants completed four trials where the memory cue appeared before the word and four where it appeared after the word. After successfully completing the practice trials, the experiment began. To avoid confusion across trials, participants completed the encoding conditions in four blocks, the order of which was fully counterbalanced across participants. Trials were blocked first according to the timing of the memory cue (before or after). Within the cue timing block, trials were further divided according to LOP (shallow or deep). For example, the order of the conditions for one-fourth of participants was: before deep, before shallow, after deep, after shallow. Each condition contained thirty-two words (half RRR), for a total of 128 words presented at encoding.

After completing all four encoding conditions, a recognition memory test containing 256 words (128 “new”) was given. Participants were instructed to respond “old” to any word that had previously been included in the experiment, regardless of the initial memory cue. They were to respond “new” only to words that had not appeared during encoding. The corresponding keystrokes (Old = V. New = N) were presented underneath each word. The screen terminated when once a decision was made, or after 10,000 ms. Finally, a fixation screen appeared for 1000 ms to signal the beginning of a new trial.

Results

Table 2 includes the recognition memory data, including hit rates for R and F items and the size of the DF effect across encoding and cue timing manipulations. A 2 (Memory Cue: remember, forget) x 2 (Cue Timing: before, after) x 2 (Level of Processing: deep, shallow) repeated measures ANOVA was conducted on the hit rates. First, the main effect of Memory Cue was significant, F (1, 93) = 48.7, p < .001, η^2 = .34. Consistent with the typical DF effect, R items were remembered better (M = .77, SE = .01) than F items (M = .68, SE = .01). A main effect of Cue Timing was also observed, F (1, 93) = 22.7, p < .001, η^2 = .20. Memory was better when the cue appeared after the word (M = .75, SE = .01), relative to before the word (M = .70, SE = .01).

Table 2

Mean proportion of items correctly identified in Experiment 1 as "old" (standard error), and the size of the directed forgetting (DF) effect, as a function of memory cue, level of processing (LOP) and cue timing.

LOP	Cue Timing	R items	F items	DF effect
Shallow	Before	.76 (.02)	.64 (.02)	0.12
	After	.77 (.02)	.71 (.02)	0.06
Deep	Before	.75 (.02)	.64 (.02)	0.11
	After	.79 (.02)	.71 (.02)	0.08

A significant Memory Cue x Cue Timing interaction, F (1, 93) = 4.69, p = .03, η^2 = .05, revealed that the size of the DF effect varied with level of Cue

Timing (Figure 2). Specifically, the DF effect was larger when the cue was presented before the word, relative to after, though planned comparisons show that the DF effect was significant in both the pre-cue and post-cue conditions ($p < .001$). Because TSS and attentional inhibition make different predictions about how memory for F items should be affected by the cue-timing manipulation, comparisons of memory cue across cue timing were also conducted. This test revealed that memory for F items was significantly lower in the pre-cue conditions, relative to the post-cue conditions ($p < .001$). The same pattern emerged for R items ($p = .01$). Finally, the main effect of LOP was not significant, F (1, 93) = .186, p = ns, and no other interactions emerged.

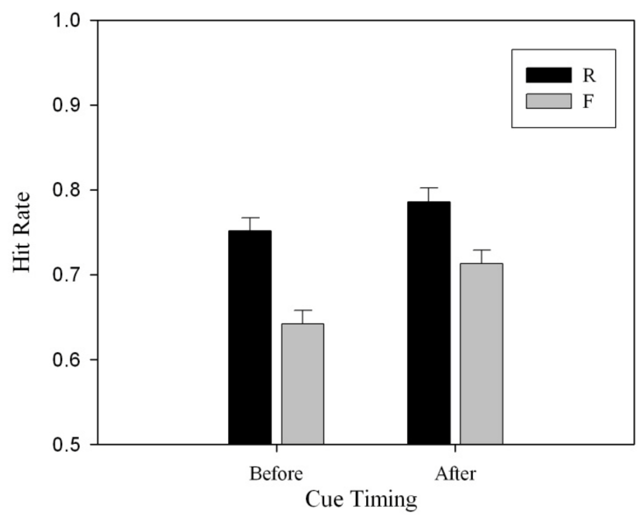

Figure 2. Mean hit rate for R and F words when the memory cue was presented before and after the target word. Error bars represent one standard error of the mean.

Discussion

Experiment 1 included two manipulations, LOP and Cue Timing, in an attempt to distinguish between two competing theories of DF: TSS and attentional inhibition. The theories make different predictions about how these manipulations should influence the size of the DF effect. Thus, interactions with these factors and Memory Cue were expected to inform current theoretical debate. Unfortunately, the evidence was inconclusive, as neither theory received unequivocal support. As expected, the classic DF effect appeared, with memory for R items surpassing memory for F items. The main effect of Cue Timing was not expected, but indicated that presenting a cue prior to the word actually hinders memory. Critical to the current predictions, a significant Memory Cue x Cue Timing interaction emerged, revealing that the size of the DF effect was smaller when a pre-cue was given, relative to a post-cue. Both theories predicted this general pattern of results, but for different reasons. TSS predicts that memory for F items should be worse with a pre-cue, relative to a post-cue, while attentional inhibition predicts no difference in F performance across pre- and post-cues. Planned comparisons revealed that memory for F items was worse with a pre-cue, relative to a post-cue, as TSS predicts. Although the pattern for F items was consistent with TSS, the pattern for R items was not. Both theories state the memory for R items should be enhanced with a pre-cue, and planned comparisons revealed that memory for R items was worse when the cue was given prior to the word. This result is discussed in greater detail below. Overall, the Memory Cue x Cue Timing interaction provides mild support for TSS, as it is consistent with the notion that a pre-cue reduces memory for F items. The LOP x Memory Cue interaction provided a stronger test of the two theories: TSS predicts that the DF effect should be larger under deep, relative to shallow encoding conditions, whereas attentional inhibition predicts the DF effect should be equal to or smaller under deep encoding, relative to shallow encoding conditions. Unfortunately, this interaction did not emerge, hampering the ability to discriminate between the two theories. Taken together, the predicted interactions were generally inconclusive, though small support for TSS came from the Memory Cue x Cue Timing results.

Before abandoning this experiment as ineffective, several other results deserve mention. A secondary goal of the current experiment was to speak to inconsistencies in the DF literature on the effect of LOP. Wetzel (1975) failed to find a main effect of LOP, but did find a LOP x Memory Cue interaction, although the pattern was opposite of that predicted for R items. Conversely, Horton and Petruk (1980) found the typical LOP effect for both R and F items, in addition to a LOP x Memory Cue interaction. The current results did not replicate either pattern of previous findings, but do speak to the previous results. First, the current results are consistent with Wetzel (1975) in showing that deep encoding does not improve (or reduce) hits to F items (Table 2). This suggested that the benefit provided via deep encoding does not disrupt the ability to forget, as hits for F items were nearly identical in the deep and shallow encoding conditions across both sets of experiments. Second, both Wetzel (1975) and Horton and Petruk (1980) found that LOP affected memory for R items, though the patterns were contradictory. In the current experiment, R items were not affected by the LOP manipulation. Although interpreting null effects is often discouraged (Keppel & Wickens, 2004), this result may have theoretical implications. If DF is exclusively dependent on selective rehearsal at encoding, then a manipulation designed to enhance rehearsal should improve performance, at least for R items. Contrary to this prediction, no benefit for R items was observed, calling into question the role that selective rehearsal plays in DF. It is of additional interest that a stable pattern of results has failed to emerge across three separate examinations of LOP in DF. Thus, it may be the case that LOP is not an informative manipulation for tests of DF.

One other result warrants mention. The main effect of Cue Timing indicates that presenting a memory cue prior to the word actually hinders memory performance. This result is contrary to the predictions of TSS and attentional inhibition, as well as the selective rehearsal theory. Why would memory be better for post-cues, relative to pre-cues? The answer may lie in the specific design of this experiment. All DF theories assume that differential processing of R and F items occurs once the memory cue is presented. In the current pre-cue conditions, this processing begins immediately, but is interrupted by the orienting task. Although it may be the case that the orienting decision is made without a cost to the primary remembering/forgetting process, the current results could be taken as evidence that this task interferes with primary processing. In other words, the LOP task may act to divide attention at encoding, which is known to hinder memory performance at testing (Pashler, 1994). Conversely, when the memory cue is presented after the word, all LOP processing has occurred. Therefore, the primary task of remembering or forgetting is left uncontaminated, which could lead to improved memory overall. Furthermore, it may be the case that the size of the DF effect is exaggerated in the pre-cue condition because this interference has differential effects on R and F items. That is, the cost of the orienting decision may be greater for F items, and smaller for R items[2].

This explanation is consistent with the current results, but cannot be confirmed given the current design. Testing this hypothesis could be done in two ways. First, a condition without the LOP manipulation but with the Cue Timing manipulation would be necessary to determine if the benefit of pre-cues emerges when a secondary task is not included. A second condition replacing the LOP decision with another divided attention task would provide evidence that dividing attention reduces memory over all, and could speak to the changes in the size of the DF effect. A divided attention task that is agnostic with respect to memory, like visual search (Pashler, 1989), should produce equivalent costs to remembering and forgetting. If a Cue Timing x Memory Cue interaction still emerges under this condition, it would suggest that divided attention is not the sole explanation for the current pattern of results.

Taken as a whole, the results do not provide strong support for either theory, although TSS did correctly predict the effect of F items across Cue Timing. Interpreting one correct prediction in the wake of many incorrect predictions seems imprudent, as the evidence should be evaluated as a whole. Thus, I am hesitant to make strong statements about either theory based on this experiment alone. It should be noted that several versions of this experiment were conducted, with modifications in design and instruction, but the results were

[2] In fact, it should be the case that R items would not be disrupted as much as F items with a LOP task, as LOP is designed to facilitate remembering. The effects of LOP on forgetting are, as such, still unknown.

consistently ambiguous. Thus, I do not believe that current inability to distinguish between the theories is a result of a strange sample or methodological deficiencies. I believe the failure to find an effect of LOP effect occurred because the effect is not present. The current results call into question the role that selective rehearsal plays in DF, but do not effectively discriminate between TSS and attentional inhibition. Experiments 2 and 3 provide additional opportunities to examine these theories under unique testing methods.

Experiment 2: DF of images, words, and faces.

As mentioned previously, Fawcett and Taylor (2008) recently compared the selective rehearsal and attentional inhibition accounts of DF. Using a probe-detection task, they found evidence that forgetting is more effortful than remembering, a finding that is inconsistent with the passive decay process described by selective rehearsal. Their result was consistent with the attentional inhibition theory of DF (Zacks et al., 1996), which states that forgetting is accomplished via the activation of an effortful inhibitory process. Interestingly, the mechanism by which attentional inhibition reduces memory for F items has been described differently by Zacks, Radvansky, and Hasher (1996) and Fawcett and Taylor (2008). Zacks et al. (1996) said "directed forgetting involves multiple mechanisms including two of an inhibitory nature…stopping the rehearsal of an item following the presentation of a forget cue, and the inhibition of TBF items at retrieval" (p. 144). Fawcett and Taylor's (2008) item probe data reflect processing at encoding, thus their explanation should be focused on rehearsal suppression. In fact, Fawcett and Taylor do not espouse rehearsal suppression, saying "attentional mechanisms…actively suppress the representation of F words at study" (p. 1169; see also Quinlan, Taylor & Fawcett, 2010). Although both explanations have adopted the term *attentional inhibition*, it is important to differentiate between the theories, as they invoke different cognitive mechanisms. Suppression of an item's representation implies more than simply ending a rehearsal process; it suggests an active attempt to remove an item from memory. In the subsequent discussion, I will use the term *rehearsal suppression* to refer to Zacks et al.'s (1996) original conceptualization, and *representation suppression* to refer to Fawcett and Taylor's (2008) modified version.

Importantly, both theories are capable of explaining the item probe results obtained by Fawcett and Taylor (2008), thus neither explanation can currently be excluded. One way to discriminate between these two theories is to include stimuli that are not amenable to maintenance rehearsal. If rehearsal is not engaged, a process to suppress rehearsal would not be needed for F items, allowing for a direct test of the rehearsal and representation suppression theories. Pictures of non-famous faces provide one such option. Faces of unknown individuals are not associated with a linguistic label that can be easily rehearsed (Ellis, 1975). Furthermore, the complex nature of faces makes verbal rehearsal of the individual parts (thin lips, big nose) overwhelming for a limited capacity phonological loop (Baddeley, 1986), especially when the duration between stimuli is short, as in an item-method DF design. Recent evidence supports the notion that faces are not verbally rehearsed. Horton, Hay and Smyth (2008; see also Smyth, Hay, Hitch & Horton, 2005) asked participants to remember the serial order of non-famous faces, either without a secondary task or while under conditions of concurrent verbal suppression (repeating the numbers 1, 2, 3, and 4), which should disrupt verbal rehearsal. Memory performance was equivalent across the two conditions, suggesting that verbal rehearsal did not contribute to memory for faces.

Currently, only one study has included DF instructions for faces, and the methodology was unorthodox. Paller et al., (1999) were interested in the neural correlates of memory for faces. Consistent with an item-method approach, they presented faces for encoding followed by an R or F cue, but three modifications made their design unique. First, fake biographical information was presented aurally at the onset of each R face to improve memory for those faces. Second, all stimuli were presented seven times during encoding, presumably to facilitate the acquisition and interpretation of EEG data. Third, they included two unique testing methods. Experiment 1 included an implicit memory task, wherein participants made famous/non-famous judgments to degraded R, F, and new stimuli. In Experiment 2, a modified recognition test was used, wherein participants were presented all stimuli simultaneously on paper, and were asked to circle the R and F faces. Paller et al. found DF effects during the modified recognition memory test (Experiment 2), but not for the implicit memory test (Experiment 1). Although the results suggest that intentional forgetting of faces can occur, the modifications make interpreting the effect difficult.

The current experiment provides the first systematic test of directed forgetting of faces. In addition, a probe task was embedded in the item-method

DF design, to provide a measure of the effort required to remember and forget. If the assumption that faces are not rehearsed is true, several findings should be observed. First, memory performance should be worse for faces, relative to other stimuli (like words or images) which are easily rehearsed. In addition, the size of the DF effect may be smaller for faces, relative to words, as R faces would not benefit from rehearsal. Finally, RTs on the item-probe task should be faster when rehearsal is not occurring, relative to conditions where it is occurring, thus RTs should be faster for faces, overall, compared to other stimuli like words or images. These outcomes are predicted by both the rehearsal suppression and the representation suppression theories of DF. In order to distinguish between them, comparisons between probe RTs to R faces and F faces must be made. According to rehearsal suppression, an active process suppresses the rehearsal of F items. If rehearsal is not engaged for faces, a suppression process would not be necessary to inhibit rehearsal after F faces. Thus, RTs for R faces should be similar to F faces. According to representation suppression, an active process should be engaged to suppress F faces, irrespective of the presence of a rehearsal process. Thus, RTs to F faces should be slowed, relative to R faces.

In addition to faces, conditions with images and words were also included. Memory for images has consistently been superior to memory for other objects, leading Shepard (1967) to coin the phrase the *picture-superiority effect*. Recent evidence suggests that that memory capacity for images is incredibly large and robust. Brady et al. (2008) presented participants with 2,500 color images over a 5.5 hour testing session. Participants were told to try and remember the details of each item, as very similar items could be shown on the memory test. At testing, a two-alternative forced choice (AFC) test was given, and participants selected the previously presented image. The foils were selected from one of three categories: novel, exemplar and state. In the novel condition, the target and foil were categorically different, for example a dresser and a train car. In the exemplar condition, the target and foil were derived from the same basic category, for example a red dresser and a black dresser. In the state condition, the target and the foil were identical objects, but the foil was in a different position or state from the original, for example a red dresser with all drawers closed and a red dresser with one drawer open. They found that participants were able to detect old objects 93%, 88%, and 87% of the time in the novel, exemplar, and state conditions respectively. This evidence suggests that images are stored with incredibly high fidelity for a long time.

Although much attention has been dedicated to memory for images, only a few studies have explored intentional attempts to forget such items. Within DF studies, stimuli including line drawings (Basden & Basden, 1996; Goernert, Widner, & Otani, 2007) and images of common objects have been examined (Quinlan, Taylor & Fawcett, 2010). One potential problem with these stimuli is that many images and line-drawings are easily verbalized. Thus people could be storing a verbal representation of the image, in addition (or instead) of the image itself. Hauswald & Kissler (2008) tried to avoid this confound by using colored scenes as stimuli in an item-method DF test. At recognition, they presented foils that were categorically matched with the old images to prevent participants from simply using a verbal term to discriminate between old and new items. The DF effect emerged, but the effect was smaller than is usually found with words, though no direct comparison with words was conducted.

Experiment 1 provides the first standardized examination of intentional forgetting across words, images and faces. Including conditions with words and images allowed for statements about the relative size of the DF effect in faces to be considered. Furthermore, the inclusion of a probe task allowed for estimates of the effort required to remember and forget to be compared across the three stimulus classes. Fawcett and Taylor (2008) observed that RTs to the probe were slower following F words, relative to R words. Both rehearsal suppression and representation suppression predict that the same pattern will emerge for images, which are also amenable to verbal rehearsal. However, if this pattern is replicated with faces, it will suggest that representations of F items are suppressed, as opposed to rehearsal being stopped.

Method

Participants

In total, 230 students enrolled in introductory psychology at Arizona State University participated for partial course credit. Each student participated in one of the following conditions: words, images, or faces. Data from fifteen participants (four from words, eight from images, and three from faces) were not analyzed, because their average RT during the item-probe procedure fell more than two standard deviations above the group mean (Winer, 1971). This left 75 participants the word condition, 66 in the image condition, and 74 in the face condition.

Materials

One hundred-twenty words, images, or faces were presented per condition. All stimuli were presented centered on a 17-in. (16-in. viewable) NEC CRT screen, with resolution set to 640 x 480. Faces and images were sized to 256 x 256 pixels, and words were presented in 18 point Calibri font in black on a white background. Images containing pictures of white, non-famous faces (half men) with neutral expressions were selected from the JACNeuF (Ekman & Matsumoto, 1993), FERET (Philips et al., 1998; Philips et al., 2000) and Minear and Park (2004) face databases. Any extraneous information that could provide an additional retrieval cue, such as background color, was removed. Words and images were matched across conditions, so that each image had a semantic associate in the word condition. To this end, high-resolution color images of single objects were first selected from the Massive Memory database (Brady, Konkle, Alvarez, & Oliva, 2008). Images were selected that contained common objects that could be easily named with a single word, and words were generated afterward to describe each image (see Appendix B).

Procedure

The experiment was conducted in a large testing room with up to nine participants per session. The experiment was programmed using E-Prime (Psychological Software Tools; Schneider et al., 2002) and presented on Dell PC computers.

Study phase

The procedure was similar to that of Fawcett and Taylor (2008), with a few notable modifications. Participants were told they would be shown either words, images, or faces for a later memory test, and each stimulus would be followed by a cue to either remember (RRR) or forget (FFF) that item. It was further explained that, because a large number of stimuli would be presented, they should try to forget the F items and only remember the R items. Sixty stimuli were presented during study, and randomly assigned with equal probability to an R or F cue at the time of encoding. In addition to the DF task, an item-probe task was administered. Following each memory cue[3], a single visual probe (a red "+") was presented, and participants responded as soon as they detected the stimulus by pressing the space key. The probe appeared with equal probability at one of three post-cue SOAs: 1,000 ms, 1,400 ms, 2,200 ms[4]. Participants were instructed that memorizing was their primary task, but to use the index finger of the dominant hand to quickly respond to the probes.

As shown in Figure 3, a single trial progressed as follows: the stimulus was presented for 2000 ms, followed by a blank screen for 1,000 ms, and then the memory cue for 1500 ms. A second blank screen provided the delay for the probe, and appeared for either 1000, 1400 or 2200 ms, followed by the probe. The probe

---

[3] Fawcett and Taylor presented the item probe either before the memory cue or after the memory cue, but only analyzed data from the post-cue probes. Because it would double the number of trials unnecessarily, pre-cue probes were not included in the current experiment.

[4] SOAs were selected to be shorter on average than those used by Fawcett and Taylor, as their results suggested that processing differences occur shortly after the memory cue has been presented.

appeared until the participant responded, or terminated at 5000 ms. A blank screen followed the probe, and appeared for either 2850 ms, 2450 ms, or 1650 ms respectively, based on the level of SOA selected for that trial. This post-probe manipulation ensured that the time allotted to process each stimulus was approximately the same across the trials, though variation in RTs to the probe added some variability. Finally a fixation stimulus (***) appeared for 1000 ms and signaled the beginning of a new trial. Overall, a single trial took 9350 ms plus the time required to respond to the probe.

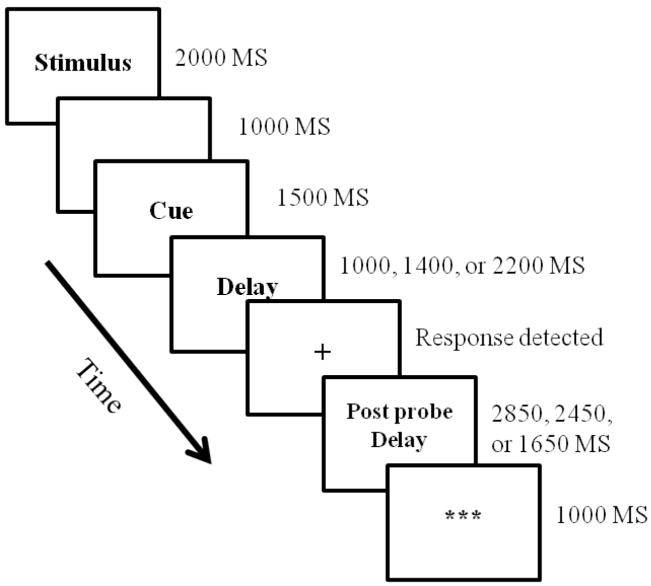

Figure 3. The timing of an individual trial in Experiment 2.

Recognition phase

At testing, 60 targets and 60 foils were presented in random order, with targets and foils counterbalanced across participants. Stimuli were presented individually and centered on a white background. Below each stimulus appeared “old” and “new” labels and the corresponding responses, “V” and “N”, respectively. Despite the instructions provided in the study phase, participants were told to ignore the previous R and F cues, and report all items from the study phase as “old”. Recognition trials contained a fixation stimulus (***) that appeared for 1000 ms, followed by the stimulus for up to 10,000 ms. The trial ended when the participant made their recognition decision, or the 10,000-ms time frame was exhausted.

Results

Recognition

Table 3 includes the relevant recognition information, including hit rates, false alarm rates, and the size of the DF effect for each stimulus class. A 2 (Memory Cue: R, F) x 3 (Stimulus: words, images, faces) mixed model ANOVA was used to analyze hit rates, that is proportion of correct “old” responses. All post-hoc comparisons were conducted with Bonferonni adjustments. First, a main effect of Memory Cue was observed, $F\ (1, 212) = 135.24$, $p < .001$, $\eta^2 = .39$. Consistent with the classic DF effect, more R items were correctly recognized as “old” (M = .85, SE = .01), relative to F items (M = .75, SE = .01). Planned comparisons confirm the DF effect was significant at each individual level of stimulus: words ($p < .001$), faces ($p < .001$) and images ($p = .001$).

Table 3

Mean proportion of items correctly identified in Experiment 2 as "old" (standard error), size of the directed forgetting (DF) effect, and false alarm rates (FAR) across memory cues and stimuli.

Stimulus	Remember	Forget	DF Effect (R - F)	F. A. R.
Words	.87 (.01)	.68 (.02)	0.18	0.14
Images	.94 (.01)	.89 (.02)	0.05	0.03
Faces	.74 (.01)	.68 (.02)	0.06	0.12

Second, a main effect of Stimulus was observed, F (2, 212) = 59.68, $p < .001$, $\eta^2 = .36$. Images were correctly identified as "old" (M = .92, SE = .01) more often than words (M = .77, SE = .03), which were correctly recognized more than faces (M = .71, SE = .01). Planned comparisons confirmed that differences between images and faces (p = .002), images and words (p < .001), and faces and words (p < .001) were all significant. False alarm rates also differed across the stimuli, F (2, 212) = 26.45, $p < .001$. Comparisons revealed that the false alarm rate was significantly lower for images, relative to faces (p < .001) and words (p < .001), but statistically equivalent across words and faces (p = ns). Finally, a Memory Cue x Stimulus interaction occurred, F (2, 212) = 28.87, $p < .001$, $\eta^2 = .21$, suggesting that the size of the DF effect varied across the stimuli. Planned comparisons revealed that the DF effect significantly differed between words and images ($p < .001$) and words and faces ($p < .001$), but not between faces and images (p = ns).

Item Probe RTs

Figure 4 shows mean RTs for R and F items at each level of SOA for the three types of stimuli. A 2 (Memory cue: R, F) x 3 (Stimulus: image, word, faces) x 3 (SOA: 1,000, 1,400, 2,200 ms) mixed model ANOVA was conducted on the RTs, with Stimulus as the sole between subject factor. A main effect of Memory Cue was observed, F (1, 212) = 21.79, $p < .001$, $\eta^2 = .09$. Consistent with Fawcett and Taylor (2008), RTs were faster following R items (M = 545.01 ms, SE = 10.8), relative to F items (M = 583.09 ms, SE = 10.55). A main effect of Stimulus was also observed, F (2, 212) = 14.83, $p < .001$, $\eta^2 = .12$. RTs were slowest for words (M = 606.20, SE = 16.68), slightly faster for images (M = 596.87, SE = 17.78), and fastest for faces (M = 489.22, SE = 16.79). Comparisons revealed significant differences between RTs for words and faces ($p < .001$) and images and faces ($p < .001$), but no difference between words and images, (p = ns). Finally, a main effect of SOA was observed, F (1, 211) = 44.83, $p < .001$, $\eta^2 = .30$, showing that RTs decreased as SOA increased. None of the potential interactions was reliable, thus comparisons examining RT differences at each level of SOA were not conducted.

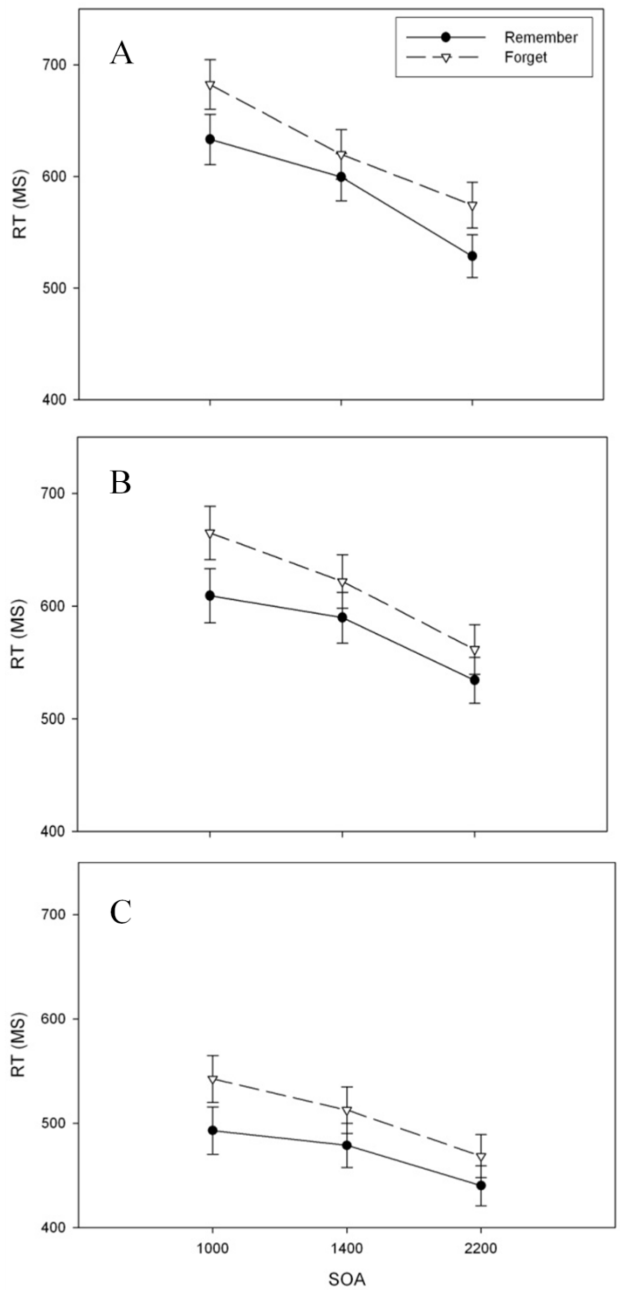

Figure 4. Mean reaction times to respond to the probe following remember or forget memory cues at each level of SOA for words (A), images (B), and faces (C). Error bars represent standard errors.

Discussion

The current experiment provided the first systematic comparison of DF across words, images and faces, and included a probe task to measure the effort expended while remembering and forgetting across stimuli. Overall, memory was worst for faces, better for words, and best for images. With respect to intentional forgetting, the classic DF effect was observed for all three stimuli, but the magnitude of the effect (R – F) varied: the DF effect was larger for words, relative to the other two stimulus classes, but was statistically equivalent across images and faces. With respect to the probe data, two findings were especially relevant. First, RTs varied across stimuli: RTs were significantly shorter for faces, relative to words and images, which were statistically equivalent. Second, examining RTs by memory cue revealed that RTs were slower following F items, relative to R items. Previous research showed this pattern in words (Fawcett & Taylor, 2008), and the current experiment extended this finding to images and faces.

The probe results provide additional evidence that forgetting in an item-method design is effortful, and not due to a passive decay process, as several theories predict. One such theory is selective rehearsal, which states that R items are rehearsed, while F items are allowed to decay (Bjork, 1972). Recently, a second passive theory, *automatic activation*, was proposed by Lee and Lee (in press). They argued that F items are ignored initially, and then reactivated through an automatic process once cognitive resources become available. Processing F items becomes more likely as time goes on because attentional resources are freed from rehearsal of R items, and are thus available to F items. Both theories predict

RTs should be as long or longer for R items, relative to F items. Automatic activation further predicts that responding to F items should be quickest at the short SOAs and either stay constant or become longer as SOA increases. The current probe data revealed that RTs to F items were longer than R items, and that RTs get faster as SOA increases. These results are wholly inconsistent with the automatic activation theory: Not only is forgetting more effortful than remembering, but the forgetting process actually becomes less effortful with time. The negative correlation between RT and SOA was also observed in the R data, which makes arguing for a tradeoff between R and F processing problematic. Thus, the current results argue against these passive theories of DF, pointing instead to an active forgetting process.

Two theories that advocate for effortful forgetting are rehearsal suppression (Zacks et al., 1996) and representation suppression (Fawcett & Taylor, 2008). Both theories state that attentional resources are required to suppress processing for F items, and both theories predicted the observed results for images and words. Faces were included because they provide a way to directly compare the two theories. According to rehearsal suppression, RTs following R and F faces should be similar, because a suppression process would not be required if rehearsal is not occurring. Conversely, the representation suppression theory suggests that an inhibitory process suppresses the representation of F items, and would be engaged even if rehearsal was not occurring. The current results showed that responding to F faces was slower, relative to R faces – a finding that is inconsistent with rehearsal suppression, and predicted by representation suppression.

Of course, espousing representation suppression over rehearsal suppression requires the acceptance of one fundamental assumption: faces are not rehearsed. Recent studies suggest that verbal rehearsal does not occur with faces (Horton, Hay & Smyth, 2008; Smyth, Hay, Hitch & Horton, 2005), and there is evidence from the current experiment that is also consistent with this notion. For example, RTs to the probe were fastest following faces, as would be expected if processing was not occurring during the post-cue delay. In addition, memory performance was worst for faces: Recognition rates were lower, relative to the other two stimuli, while false alarm rates were high. Again, this pattern would be expected if rehearsal was unable to provide a memory benefit for these items. Although not conclusive, the current findings suggest that verbal rehearsal of faces was unlikely to occur under the given conditions. When combined with previous evidence (Horton et al., 2008) it is not imprudent to assume that the current statements about DF are supported.

The results also align with a recent DF study that included a stop-signal task (Fawcett & Taylor, 2010). In the stop-signal task, participants are given a “go-signal” and asked to quickly provide a target response. On a portion of trials, a “stop-signal” is presented after the “go-signal”, indicating that the target response should be withheld. In other words, on the “stop-signal” trials, participants must suppress the response they were preparing to execute. Performance is measured by combining RTs on go-trials to the probability of inhibiting a response on stop-trials (stop-signal reaction times; SSRTs). Fawcett and Taylor wondered if the mechanism responsible for forgetting might be similar to the mechanism that is engaged in a stop-signal task. If these mechanisms are similar, then performance should be better on stop-trials that follow an F instruction (when a stop process would be engaged), relative to trials following an R instruction (where a stop process would not be engaged). That is, SSRTs should be faster following F trials, relative to R trials. They found that an F instruction increased the probability that participants would inhibit the target response, but led to longer SSRTs. They interpret this result by saying that forgetting is a distinct process (or activates distinct mechanisms) from the mechanisms responsible for stopping an ongoing process. Although Fawcett and Taylor did not interpret their results with respect to rehearsal suppression, the parallels are obvious. According to rehearsal suppression, forgetting is achieved by stopping the ongoing rehearsal process, just as the stop-signal task requires. Thus, their evidence aligns with the current findings and indicates that the active mechanism that contributes to DF is more than simply rehearsal suppression.

One final result merits further discussion: The size of the DF effect varied across the three stimuli. The effect was smaller for faces, relative to words, which would be expected if rehearsal was unable to benefit memory for R faces. Interestingly, the size of the DF effect was statistically equivalent across faces and images, although the patterns of data differed markedly between the two stimuli (Table 4). Consistent with the picture superiority effect (Shepard, 1967; Standing, 1973), images were well remembered, producing the highest recognition rates and lowest false alarm rates of the three stimulus classes. This is in stark contrast to the pattern just described for faces. One potential explanation is that the small DF effects observed for images and faces occur for different reasons: faces are not well remembered, while images are not easily forgotten. The finding that DF effects vary across stimuli could have important implications for those studying directed forgetting, as well as those studying forgetting in areas outside of cognition. For example, intentional forgetting has been studied in social psychology through work on social judgment (Johnson, 1994). In these experiments, participants are given complex scenarios and asked to provide a judgment about a person, event, or decision occurring in that scenario. Importantly, they are instructed to disregard, or forget, some piece(s) of evidence while making their decision. Forgetting is estimated by comparing the judgments of a group that did not receive the pertinent information to the group that was told to disregard such information. A real-world analog would be jury trials in which a judge might instruct the jurors to disregard physical evidence, pictures, or statements in their deliberation. The current results suggest that forgetting visual imagery, such as photographs of objects, may be less likely in these situations, and could influence subsequent judgments.

In conclusion, the results provide additional evidence that intentional forgetting in an item-method design does occur across a variety of stimuli, this process is effortful, and it is unlikely to be due to a passive decay process, as many have suggested (Basden, et al.1993). Furthermore, evidence from the face condition suggests that this effortful response is more than simply rehearsal suppression, and may be due to suppression of item representations.

Experiment 3

According to TSS, intentional forgetting is “due primarily to the discrimination of instruction sets after they have already been stored” (MacLeod, 1975, p 271). That is, R and F items are tagged with their associated memory cues and organized at encoding, and at testing a selective search process targets R items while avoiding F items. As mentioned previously, other factors including selective rehearsal of R items may contribute to performance via improved differentiation of R and F items (Bjork, 1970), but the primary mechanism is a discriminatory search process that seeks out R items and ignores F items. TSS is thus an ironic theory of intentional forgetting, in that it implies that items are not actually forgotten, they are simply tagged and avoided. Consistent with this theory, Epstein and Wilder (1972) found that recall of F items improved when participants were encouraged to search for these items. They presented word-pairs at encoding that were either TBR or TBF. At testing, one word was presented and participants had to provide the matched pair. Importantly, participants were occasionally told at testing whether the target item belong to an F pair or R pair. Consistent with the DF effect, performance was better overall for R items, relative to F items, but the size of the effect was reduced when clues about the pairs were provided. Specifically, performance improved for F items when participants were told the target item was an F item, but performance for R items remained stable. Furthermore, participants rarely recalled F items unless they were primed to search through F pairs. These results are consistent with the notion that F items are stored with their corresponding labels, and can be retrieved if appropriately prompted.

Experiment 3 examined the TSS theory of DF by including conditions that explicitly required tagging at encoding and a thorough search of all stimuli at testing. These conditions were compared to DF conditions with equivalent manipulations, to determine if tagging and selective search mechanisms contribute to intentional forgetting. The predictions were simple: parallels across conditions would be interpreted as evidence for TSS, while inconsistencies across conditions would suggest that TSS is not a sufficient explanation for intentional forgetting. With respect to encoding, conditions with DF instructions were compared to conditions where items were tagged with an arbitrary label. Across both conditions, words were presented individually and followed by a memory cue, consistent with an item-method design. In the DF condition, the memory cue was either R of F, which indicated the appropriate mnemonic process. In the *arbitrary condition*, each word was followed by an A or B cue, and participants will told to remember both sets of words and their corresponding labels. Thus, the arbitrary condition encouraged participants to tag and organize items at encoding. To examine processing at encoding, a probe-detection task (as in Experiment 2) was embedded in both the DF and Arbitrary conditions. As mentioned, RTs to the probe are taken as an estimate of the effort required by the primary task. If tagging occurs at encoding, then RTs patterns should be similar across DF and Arbitrary conditions.

Of course, tagging is only one part of the TSS theory, and selective search at testing is thought to be the primary cause of the DF effect (MacLeod, 1975). According to TSS, R items are prioritized at search, leaving F items to be underreported, despite their presence in memory. It has been suggested that recognition and recall tests encourage this discrepancy, because participants have reduced motivation to search for items in these testing designs (MacLeod, 1999). In the current experiment, recognition memory tests were included for two reasons. First, these tests have become the standard in the item-method DF literature. Second, Experiments 1 and 2 both included these tests. Thus, including recognition tests in Experiment 3 allowed for comparisons across the current experiments and a significant portion of the DF literature. To determine whether F items were actually retained but underreported, Experiment 3 included two additional testing conditions that required a thorough memory search for successful performance: process dissociation procedure and source monitoring. Across both methods, memory that an item was presented previously was insufficient to make a correct decision, and information about the memory cue associated with the word was also necessary. If F items are stored but underreported, then explicit instructions to search memory should increase retrieval of F items (Epstein & Wilder, 1972).

Process Dissociation Procedure

Jacoby's (1991) process dissociation procedure (PDP) provides an intriguing testing option for several reasons. First, PDP requires participants to engage in a thorough search of memory via unique testing instructions (described below). Furthermore, PDP can be used to estimate the contributions of recollection and familiarity to overall memory performance, providing an additional set of dependent measures to explore. Finally, PDP has been used extensively in memory research, but not yet applied to DF. In one early example of the PDP paradigm (Jacoby, 1991, Experiment 3), two lists of words were presented, one visually and one aurally. At recognition, participants responded "old" or "new" to targets and lures. In the *inclusion* condition, participants were asked to respond "old" if the word was presented in either list. However, in the *exclusion* condition, participants only responded "old" to items in the auditory list. Under inclusion instructions, both recollection and familiarity can be used to facilitate recognition for words on both the visual and auditory lists, because specific details regarding source are not necessary. Conversely, source details are necessary for successful performance in the exclusion condition. In this condition, "old" responses to words from the auditory list may be generated from recollection or familiarity, but "old" responses to words from the visual list would arise from familiarity, coupled with a failure of source recollection. Critically, an estimate of recollection can be derived from these two conditions. As Jacoby (1991) writes,

> "… recollection can be measured as the difference between the likelihood of responding to an item of a given class when people are attempting to select *for* items of that class as compared to when they are attempting to select *against* items of that class." (pg 526)

Thus, by comparing performance in the Inclusion and Exclusion conditions, the contribution of recollection can be estimated, and from there, estimates of familiarity can be calculated.

The current experiment included Inclusion and Exclusion instructions for both the DF and Arbitrary conditions, so that estimates of recollection and familiarity could be obtained for each condition and compared. In typical DF experiments, participants are asked to respond “old” to every word that was previously presented, regardless of the initial R/F cue. This is essentially the equivalent of Jacoby’s inclusion condition, where R and F source details are unnecessary for accurate performance. As far as I am aware, no DF studies have yet to include the equivalent of an exclusion condition. In the current experiment, exclusion instructions required participants to call R items “old” and F items “new”. In the Arbitrary Inclusion condition, participants were to respond “old” to both A and B items, while in the Arbitrary Exclusion condition “old” responses were only made to A items. Thus, the instructions for F and B items change across these two conditions. Mirroring the procedure used by Jacoby, the probabilities that F and B items were called “old” in the inclusion and exclusion conditions were used to estimate the contributions of recollection and familiarity across the DF and Arbitrary condition. If TSS is correct, and details about memory cues are stored, then estimates of recollection and familiarity should be equivalent across the DF and Arbitrary conditions. Alternatively, if TSS is not correct, then these values may be different.

Source Monitoring

Source monitoring is similar to PDP in that episodic details are required for accurate memory performance, but it differs from PDP in that the responses are explicit. Unlike typical recall and recognition tests, source monitoring tests require participants to make an additional discrimination about details of the stimuli that were presented at encoding. These details could include the color in which a word was presented, the name or gender of the individual providing information, the side of a screen on which a visual stimulus appeared, or the list to which an item belonged (Johnson, Hashtroudi, & Lindsay, 1993). In the current experiment, *source* referred to the memory cue that was associated with the item at encoding. During recognition, participants were given a three-alternative forced choice test and asked to identify each item as new, old-source1 or old-source2. As in PDP, the instructions encouraged participants to search for all relevant information in order to make an appropriate decision.

Source monitoring has been included previously in studies of intentional forgetting. For example, MacLeod (1975) presented DF instructions using the item method and gave participants a modified recognition test either one or two weeks after encoding. At testing, he provided a list containing old and new words, and asked participants to circle all previously shown words. Half of the words were presented with their corresponding memory cues, but participants were asked to fill in the correct cues for the remaining half. He found that memory for source was better than chance after both the one and two week delays, suggesting that memory for cue information was relatively robust. He also found that source memory was better for R items, relative to F items. That is, among items that were correctly called “old”, R items were correctly classified as R more often than F words were identified as F.

The current experiment extends the use of source monitoring by including an analytic technique that has been underutilized in the DF literature. Multinomial modeling is ideal for analyzing categorical or source based data, because it provides a method for separating the contribution of item discrimination (old or new), source detection (source 1 or source 2), perceiver bias, and guessing to performance. As outlined by Batchelder and Riefer (1990), multinomial modeling utilizes processing trees that reflect the number of sources, the assumed psychological processes, and the type of items presented at testing for a given experimental design (see Figure 5). The term “tree” reflects the physical structure these models assume. Beginning with the superordinate psychological state (most often detection), a series of branches emerge that contain conditional link probabilities to the other states or the final response. Each link is characterized by either the presence of a parameter (P) or the inverse of that parameter (1 – P). Thus, a series of equations can be derived that lead from the starting state to the final response, simply by following the paths and multiplying the relevant equations. Finally, estimates for the parameters are derived from summed response frequencies, and maximum likelihood estimation is used to solve across the multiplied equations. The resulting parameter estimates reflect the relative contribution of each psychological state to performance in the form of probabilities.

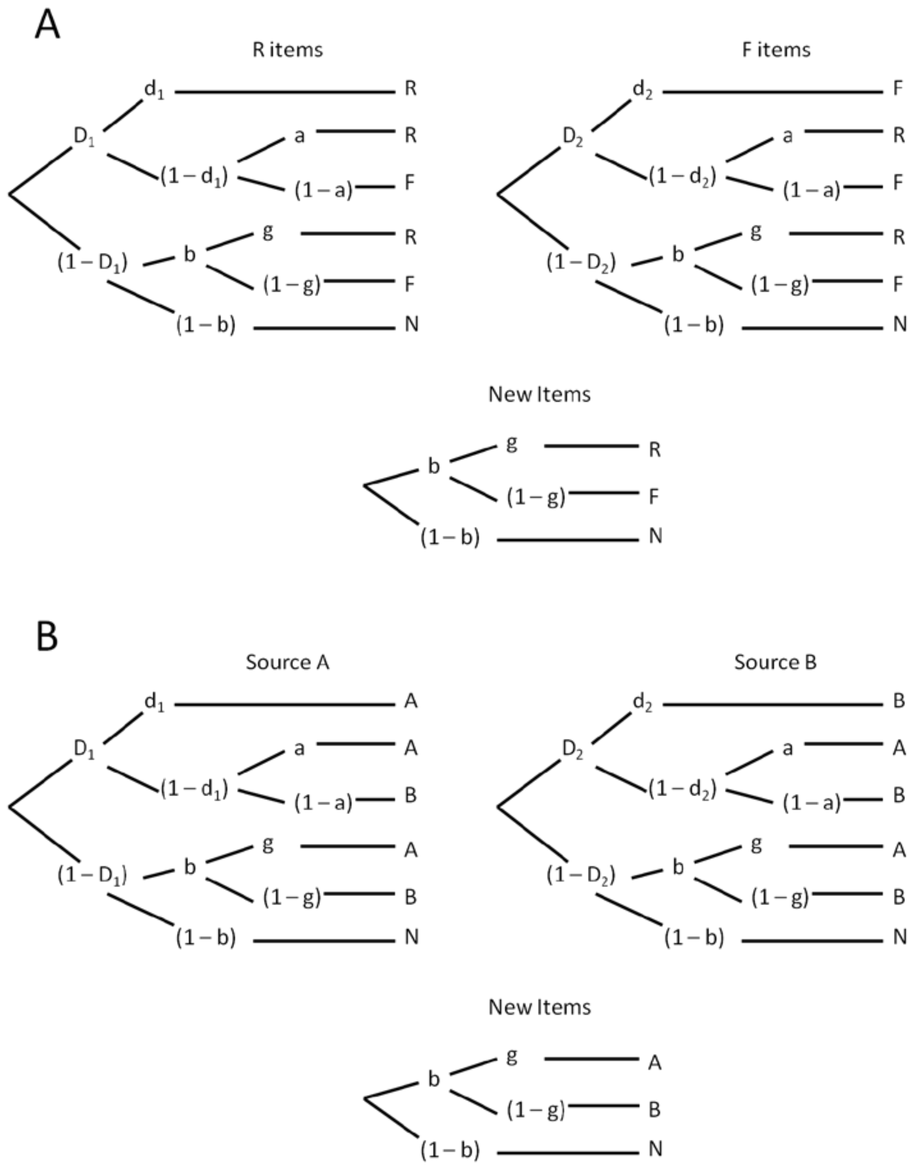

Figure 5. The seven parameter multinomial processing tree for the DF (A) and Arbitrary (B) Conditions. (D_1 = detectability of Source R or A items; D_2 = detectability of source F or B items; d_1 = source discrimination for the R or A items; d_2 = source discrimination for the F or B items; b = bias for responding "old" to a non-detected item; a = guessing that a non-discriminated item belongs to Source R or A; g = guessing that a non-detected item belongs to Source R or A.)

Multinomial modeling provides a helpful tool in examining intentional forgetting because it allows for the contributions of various processes to be estimated independently for remember and forget trials. Applying this technique is likely to improve our understanding of the processes involved in forgetting, and provides another way to test the various theories of DF. If TSS is correct, then source discrimination for both R and F items should be relatively good, as the theory states that the items are tagged and organized separately. Currently only one examination of DF has included multinomial modeling, and that was conducted on the list method. Sahakyan and Delaney (2005) compared groups of participants that were asked to forget one set of items (List 1) and remember a second set (List 2) to a group that was required to remember both List 1 and List 2. Participants were asked to recall all items they could, and name the list on which each item appeared. They found that the forget group erroneously identified List 1 items as List 2 items, while the remember group often identified List 2 items as List 1 items. Multinomial modeling was conducted to further examine the nature of this interaction, and revealed that source detection did not differ across groups, but guessing did. They found that the forget group was biased to guess that detected items appeared in List 2 (the TBR list), whereas the remember group was biased to guess that detected items appeared in List 1. Their results suggest that multinomial modeling can be a useful tool in distinguishing between contributing factors in intentional forgetting. The current experiment extends their work by including the first application of multinomial modeling to item method DF.

In summary, the current experiment examines the TSS theory by including encoding and testing instructions that encourage participants to adopt a TSS approach. Figure 6 provides a schematic overview of the tests, memory cues, and the conditions that were included in Experiment 3. Three testing procedures were included: Recognition memory, process dissociation procedure and source monitoring. Within each testing method, two encoding conditions were compared, Arbitrary and DF, which differed in the nature of the memory cue. A control condition was also included, during which participants had to remember all items and no memory cues were provided (Words only). This condition allowed for assessments of the relative memory performance in the DF and Arbitrary conditions to be made. If TSS is correct, then the presence of memory cues in the DF and Arbitrary condition may double the memory load, which would lead to reduced performance in these conditions, relative to a Words Only condition. On the other hand, if TSS is incorrect, and tagging occurs in the Arbitrary condition only, then performance is expected to vary across the Arbitrary, DF and Words Only conditions. Performance should be better in the Words Only condition, relative to the Arbitrary condition, as just described. Additionally, if forgetting does occur, then memory performance should be better for R items and worse for F items, relative to the Words Only condition.

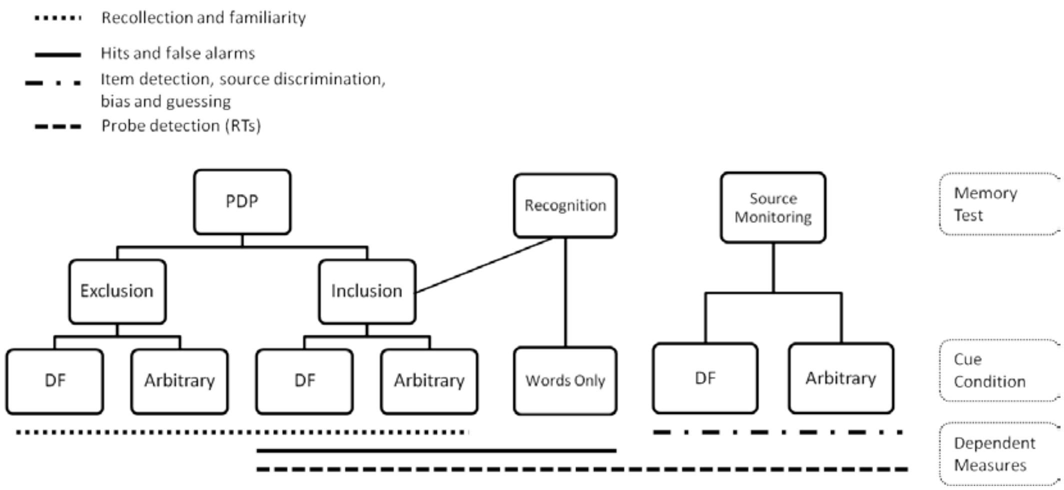

Figure 6. A schematic overview of Experiment 3 including the three memory tests, seven experimental conditions, planned comparisons across conditions, and critical dependent measures.

In total, seven between-subject conditions were included to fully inspect the TSS theory: DF Exclusion, DF Inclusion, Arbitrary Exclusion, Arbitrary inclusion, Remember All, DF Source Monitoring, and Arbitrary Source Monitoring. The lines underneath the seven conditions in Figure 6 illustrate the comparisons that were made across the various conditions and dependent measures. A probe detection task was embedded within each condition, which was included to provide an estimate of effort required at encoding. In addition, several dependent memory measures were included. Estimates of Recollection and Familiarity were derived from the inclusion and exclusion conditions. Typical measures of recognition memory, like hits and false alarms, were included when appropriate. Finally, multinomial modeling was conducted on the source monitoring conditions to estimate the contribution to a variety of cognitive processes to directed forgetting. One basic prediction tied these assorted manipulations together: If TSS is the mechanism used in DF, patterns of performance should be similar for DF and Arbitrary conditions across these manipulations and dependent measures.

Method

Participants

A total of 346 introductory psychology students at Arizona State University participated in the experiment for partial course credit. Each participant completed one of the seven conditions. Data from 41 participants were not included either because recognition accuracy fell below chance (33% in source monitoring conditions, 50% in all other conditions) or because average RT to the probe-detection task was more than two standard deviations from the mean (Winer, 1971). This left data from 305 participants for final analysis: 43 in DF Exclusion, 41 in DF Inclusion, 46 in Arbitrary Exclusion, 37 in Arbitrary Inclusion, 44 in Words Only, 49 in DF source monitoring, and 45 in Arbitrary source monitoring.

Materials

One-hundred twenty words were selected from the MRC database (Wilson, 1988; see Appendix C). Words were selected that contained between one and three syllables and scored higher than 500 (out of 700) on a scale of familiarity. Two sets of 60 words were created prior to conducting the experiment to serve as an encoding list and foils list for the recognition test. Words were not matched across lists for factors such as word length or frequency, but two version of each condition were created so that targets and foils were counterbalanced across participants. Words on the encoding list were randomly assigned with equal probability to a memory cue at the beginning of the experiment. Words were presented in capital letters in 18 point Courier New font in black on a gray background. The probe, a “+”, appeared in 18 point Courier New font in red.

Procedure

The experiment was run in a large testing room with up to ten participants per session. The experiment was designed using E-prime software (Psychological Software Tools; Schneider, et al., 2002) and presented on Dell PC computers with a 17-in. (16-in. viewable) CRT computer screen.

Study Phase

The study phase was similar to Experiment 2, with a few modifications. Participants were told they would be shown words followed by memory cues. In the DF condition, the memory cue was either RRR or FFF, which also signaled the mnemonic process that should be engaged. In the arbitrary condition, the cue was AAA or BBB and participants were instructed to remember all words and their corresponding cues. In the Words Only condition, no memory cue was presented, and participants were simply instructed to memorize every word. The probe-detection task matched Experiment 2 exactly.

With the exception of the Words Only condition, the basic encoding procedure was the same across conditions. Figure 7 illustrates a typical trial: A word appeared for 1000 ms, followed by a blank screen appearing for 1000 ms, and then the memory cue for 500 ms. After a delay of 1000, 1400, or 2200 ms, which reflected the level of SOA for that trial, the probe appeared and remained on the screen until a response was detected or 5000 ms had elapsed. Following the probe, a blank screen appeared for 2850, 2450, or 1650 ms depending on the level of SOA. Finally, a fixation screen (***) was presented for 1000 ms and signaled the beginning of a new trial. Overall, an individual trial took 7350 ms plus the time required to respond to the probe.

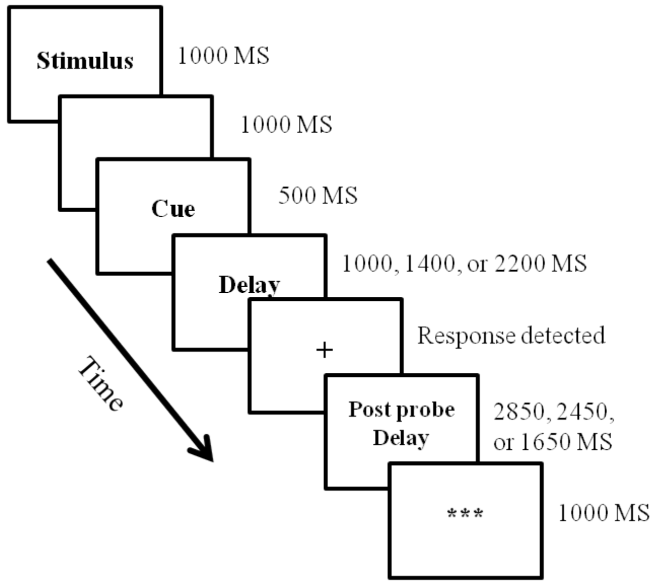

Figure 7. An example of the timing of a typical encoding trial for the DF Exclusion, DF Inclusion, Arbitrary Exclusion, Arbitrary Inclusion, DF Source Monitoring and Arbitrary Source Monitoring conditions. During the Words Only condition, the post-stimulus delay and memory cue were not presented.

In the Words Only condition, no memory cues were presented. To ensure that SOA timing was consistent across conditions, the post-stimulus blank screen was also removed. Thus, the procedure for this condition included the stimulus, SOA delay, probe, post-probe delay and the fixation screen. The timing for an individual trial was 5850 ms plus the time required to respond to the probe. Although the trial was shorter overall, the rehearsal time was nearly equivalent to the other conditions (approximately 4850 ms plus time to respond to the probe), as rehearsal in the Arbitrary and DF conditions was not expected to begin until after the memory cue appeared.

Recognition

During the testing phase, 60 targets and 60 foils were presented individually in a random order. All recognition trials contained a fixation stimulus (***) that appeared for 1000 ms, followed by a screen containing the stimulus and all possible responses. The display and testing instructions varied across conditions. In the exclusion conditions, participants responded “old” only to words that were associated with an RRR label in the DF condition or the AAA label in the arbitrary condition. Items that were associated with FFF or BBB cues were to be called “new” in addition to any foils. In the inclusion conditions, all words that had been previously shown, regardless of the associated memory cue, were to be called “old”, and only foils were to be called “new”. Similarly, in the Words Only condition, all previously presented words were to be called “old”, and foils were called “new”. Across these conditions, participants responded by pressing “V” to old items, and “N” to new items. In the source monitoring conditions, participants were given a three-alternative forced choice test, in which they had to identify the cue associated with each old word. For words paired with RRR or AAA cues, the correct response was “V”; for words associated with FFF or BBB cues, the correct response was “B”; if the word was new, the correct response was “N”. Across all conditions, the trial ended when the participant made their recognition decision, or a 10,000 ms time frame was exhausted.

Results

Inclusion and exclusion conditions

The exclusion and inclusion conditions were included so that estimates of recollection and familiarity could be derived for the DF and Arbitrary conditions. Following Jacoby's procedure, the probability of calling an F or B item "old" in the Inclusion ($P_{inclusion}$) and Exclusion ($P_{exclusion}$) condition was calculated, and are presented in Table 4[5]. Estimates of recollection and familiarity were then calculated, and are also presented in Table 4.

Table 4

Averaged probabilities that items were called "old" in the inclusion and exclusion conditions, and estimates of recollection and familiarity derived from these probabilities.

	Inclusion	Exclusion	Recollection	Familiarity
DF	0.54	.28	0.26	0.38
Arbitrary	0.74	0.45	0.29	0.63

Note: DF = directed forgetting

According to Jacoby (1992), recollection reflects source memory, and is calculated by subtracting the probability that F or B items were incorrectly called old in Exclusion from the probability these items were correctly called old in Inclusion [$P_{Inclusion} - P_{Exclusion}$]. The resulting value is a probability that ranges from 0.0 – 1.0. A value of 1.0 would indicate that F and B items were always correctly called "old" in the inclusion condition, and never incorrectly called

[5] The estimates of recollection and familiarity are based solely on memory performance from items that had to be avoided in the Exclusion condition (F or B). Therefore, memory performance for R and A items had no influence on the resulting estimates.

“old” in exclusion. Such performance would only be possible if item and source information were retained perfectly. The current values are well below 1, but above 0, indicating that source memory contributed to performance in both the DF and Arbitrary conditions.

On the other hand, familiarity reflects memory for an item, in the absence of source information. It is calculated by dividing the probability that an item was incorrectly called “old” in the exclusion condition by the probability that it was not recollected [$P_{Exclusion}$ / (1 – R)]. Familiarity estimates can also be interpreted as probabilities, and in cases when recollection is near 0, but overall memory performance is good, estimates of familiarity will be high. However, estimates of recollection and familiarity need not sum to 1, and are not always inversely correlated. Jacoby (1991) argued that recollection and familiarity are independent processes, therefore it would be possible to influence the contribution of recollection while leaving the value of familiarity unchanged, and vice versa[6]. The estimates of Familiarity presented in Table 4 indicate that Familiarity contributes to performance in both conditions, though not equally.

As Inclusion and Exclusion conditions were between subjects, $P_{Inclusion}$ and $P_{Exclusion}$ values were averaged across all participants in a condition, and single estimates of Recollection and Familiarity were calculated. Because single recollection and familiarity values were obtained, inferential statistics were not possible, but the relative values are informative. First, the estimates of recollection and familiarity were lower in the DF condition, relative to the

[6] This independence assumption is, however, controversial. See Yonelinas (2002) for a review.

Arbitrary condition. Furthermore, the patterns across DF and Arbitrary conditions differed. While the relative contribution of recollection was similar, the contribution of familiarity was quite different, and familiarity contributed more to memory in the AB condition than the DF condition. If the mechanisms contributing to performance were similar across AB and DF conditions, one would expect these patterns to be similar, but this was not the case.

The obtained estimates of recollection and familiarity argue against the TSS theory of DF, but additional analyses are warranted. Typically, hits to R and F items are examined to derive an estimate of the size of the DF effect. However, a *hit* is difficult to interpret in the exclusion conditions, given the unique instructions for F and B items. Typically, a hit is considered correctly identifying an old item as old. In the Exclusion condition, F and B items were old, but were to be labeled new. Furthermore, there are multiple ways by which participants could arrive at a “New” decision in the Exclusion condition. F and B items could be remembered with the appropriate source information, so that participants were able to use source memory to exclude these items when asked to do so. Jacoby assumed that participants used this strategy. However, these items could also be called “new” because they were forgotten, and participants believed they were new. Given that the DF condition encouraged participants to forget these items, Jacoby’s original assumption may not hold. Thus, it is not possible to determine if participants called these items “new” intentionally or unintentionally given the current conditions, which makes calculating a hit rate in the service of obtaining a DF effect problematic. Because the instructions and the implications of the exclusion conditions are quite different from all the other conditions, comparisons across these conditions are not warranted. Therefore, data from the exclusion conditions was not included in the subsequent analyses, and comparisons across DF Inclusion, Arbitrary Inclusion and the Words Only control condition were conducted.

Recognition Memory

Memory performance was compared by examining hits and false alarms across the three Cue conditions (Arbitrary, DF, Words Only). Both the Arbitrary and DF conditions contained two sets of memory cues, but the Words Only condition had none. To ensure that comparisons across conditions were based on equal numbers of trials, a set of 30 old words was selected from the Words Only condition and used for comparison with the R/A items, and the remaining 30 items were compared to F/B items. These items were randomly assigned to one of the two sets at encoding, using the same randomization procedure that was used to assign memory cues in the Arbitrary and DF conditions.

Table 5 contains the recognition data, including hit rate, false alarm rates (FAR), and corrected hit rate (Hits – FAR). A 2 (Memory Cue) x 3 (Cue Condition) mixed model ANOVA was conducted on the unadjusted hit rates. Cue condition was the only between subject factor. A main effect of Cue Condition appeared, F (2, 118) = 3.17, p = .05, η^2 = .05. Post-hoc comparisons with Bonferonni adjustments revealed that performance in the DF condition was marginally lower, relative to the Arbitrary condition (p = .06), but no other significant differences were observed. A one-way ANOVA compared false alarm rates across the three Cue Conditions, and revealed a significant effect of Cue Condition, F (2, 118) = 20.25, $p < .001$. Comparisons showed that false alarms were significantly higher in the Arbitrary, relative to the DF and Words Only conditions ($p < .001$). No difference between the DF and Words Only conditions emerged (p = ns).

Table 5

Averaged recognition memory data (standard error) as a function of cue conditions and memory cue.

Condition	Memory Cue	Hit Rate	False Alarm rate	Corrected Hit Rate
DF	R	0.82 (.02)	0.21 (.02)	0.61
	F	0.54 (.03)		0.33
Arbitrary	A	0.74 (.02)	0.46 (.04)	0.28
	B	0.75 (.03)		0.29
Words Only	Set 1	0.72 (.02)	0.23 (.02)	0.49
	Set 2	0.74 (.03)		0.51

Note: DF = directed forgetting; Corrected hit rate = (Hit Rate – False Alarm Rate)

The main effect of Memory Cue was also significant, F (1, 118) = 33.42, $p < .001$, $\eta^2 = .22$. Given that Memory Cue is interpreted differently across the three conditions, this effect is best understood in terms of the significant Memory Cue x Cue Condition interaction, F (1, 118) = 42.09, $p < .001$, $\eta^2 = .42$. Planned comparisons revealed that memory cue had no effect on recognition in the Words Only or Arbitrary conditions. However, memory cue did have an effect in the DF condition. Consistent with the typical DF effect, R items were remembered better than F items ($p < .001$). An additional set of comparisons was conducted, to determine if performance varied across the paired memory cues. Post-hoc comparisons with Bonferonni adjustments revealed that R items were remembered better than A items ($p = .026$) and the equivalent set of Words Only items ($p = .002$). Conversely, F items were not remembered as well as B ($p < .001$) or the equivalent set of Words Only items ($p < .001$). Finally, performance was statistically equivalent across Words Only and Arbitrary conditions (all comparisons p = ns). Combined, this set of comparisons suggests that forgetting did occur, as memory for F items was significantly lower in the DF condition, relative to the analogous item sets in the Arbitrary and Words Only conditions. Furthermore, a benefit was observed for R items, relative to the other conditions – a finding that is consistent with previous evidence (Muther, 1965) showing that forgetting some items improves memory for other items.

RTs to the probe-detection task provided another dependent measure of interest, albeit one that reflects processing at encoding. A 2 (Memory Cue) x 3 (Cue Condition) x 3 (SOA) mixed-model ANOVA was conducted on RTs to detect the probe, and Cue Condition was the only between subject factor. The main effect of Cue Condition was significant, $F(2, 118) = 31.38$, $p < .001$, $\eta^2 = .35$. RTs were fastest in the Words Only condition (M = 357, SE = 20), followed by the Arbitrary Condition (M = 445, SE = 21), and then the DF condition (M = 579, SE = 20). The main effect of Memory Cue was significant, $F(1, 118) = 5.5$, $p = .02$, $\eta^2 = .05$, but should be interpreted with respect to the Memory Cue x Cue Condition interaction, $F(2, 118) = 4.45$, $p = .01$, $\eta^2 = .07$ (Figure 8). Planned comparisons revealed that Memory Cue had no effect in the Arbitrary or Words Only conditions, but did have an effect in the DF condition, where RTs were slower to F items, relative to R items ($p < .001$). Finally, the main effect of SOA, F (2, 117) = 37.32, $p < .001$, $\eta^2 = .39$ and the SOA x Cue Condition interaction were both significant, F (4, 236) = 3.4, $p = .01$, $\eta^2 = .05$. These effects show that RTs decreased as SOA increased, but the magnitude of the decrease was larger in the DF condition, relative to the other conditions.

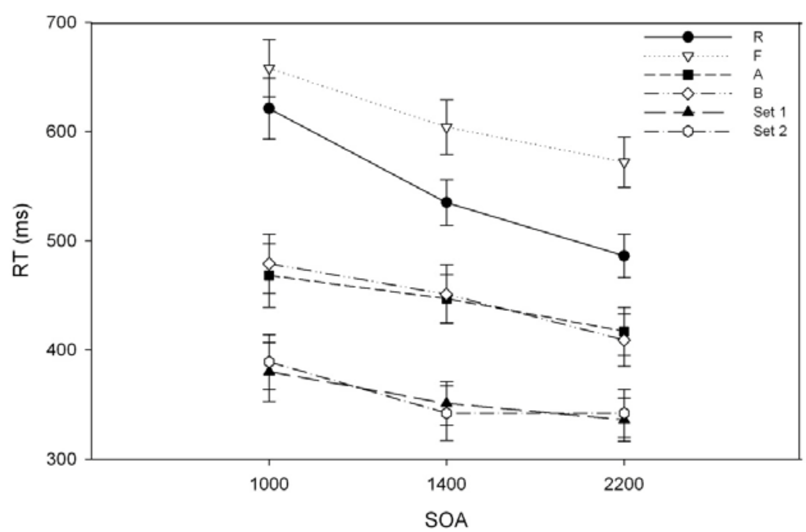

Figure 8. Average response times to detect the probe for each memory cue at each level of SOA for Inclusion conditions. Error bars represent standard errors. Set 1 and Set 2 reflect data from the Word Only condition; R and F from the directed forgetting condition; A and B from the Arbitrary condition.

Source monitoring

In the source monitoring conditions, participants had to determine the cue that was associated with each word at encoding. Two conditions, DF and Arbitrary, were included, and the memory data are presented in Table 6. A 2 (Memory Cue) x 2 (Cue Condition) mixed-model ANOVA was conducted on hit rates (calling and R item “R”, and an F item “F”), and a one-Way ANOVA was conducted on false alarm rates. Cue condition was the only between subject factor. The main effect of Cue Condition was not significant for hits, (F = 1.42, p = ns), but was significant for false alarms, F (1, 92) = 4.9, p = .03. False alarms were higher in the Arbitrary condition, relative to the DF condition. The main effect of Memory Cue emerged, F (1, 92) = 22.11, p < .001, η^2 = .19, and the Memory Cue x Cue Condition interaction was significant, F (1, 92) = 22.11, p < .001, η^2 = .19. Post-hoc comparisons revealed that memory cue had no effect in the arbitrary condition (p = ns). On the other hand, an effect of memory cue was apparent in the DF condition, as R items were identified better, relative to F items (p < .001).

Table 6

Averaged recognition memory data (standard error) from the source monitoring conditions as a function of cue conditions and memory cue.

Condition	Memory Cue	Hit Rate	False Alarm rate	Corrected Hit Rate
DF	R	0.58 (.02)	0.28 (.02)	0.3
	F	0.44 (.02)		0.16
Arbitrary	A	0.49 (.02)	0.37 (.03)	0.12
	B	0.47 (.02)		0.1

Note: DF = directed forgetting; Corrected hit rate = (Hit Rate – False Alarm Rate)

RTs to the probe detection task were analyzed in a 2 (Memory Cue) x 2 (Cue Condition) x 3 (SOA) mixed-model ANOVA (Figure 9). The main effect of Cue Condition was significant, F (1, 92) = 34.04, p < .001, η^2 = .27, and showed that RTs were faster in the Arbitrary Condition, relative to the DF condition. The effect of Memory Cue was significant, F (1, 92) = 17.24, p < .001, η^2 = .16, as was the Memory Cue x Cue Condition interaction, F (1, 92) = 18.85, p < .001, η^2 = .17. Planned comparisons revealed that there was no effect of memory cue in the Arbitrary condition, but RTs were slower following F items, relative to R items (p < .001) in the DF condition. The main effect of SOA was significant, F (2, 91) = 18.17, p < .001, η^2 = .29, indicating that RTs decreased as SOA increased, but the SOA x Condition interaction was not (F = .342; p = ns).

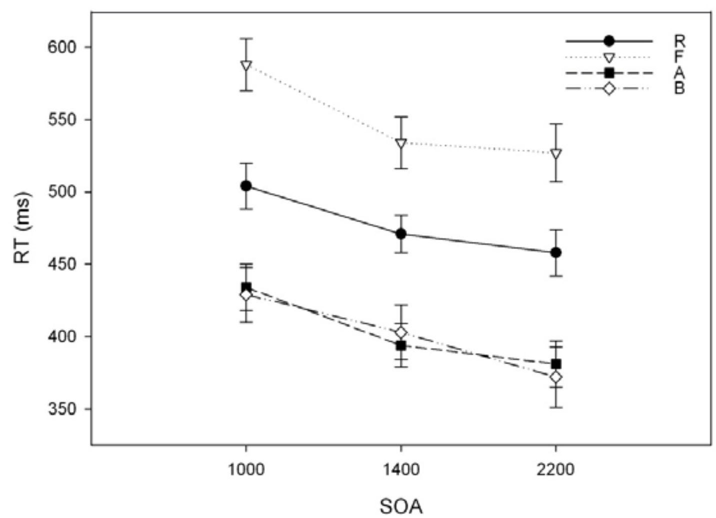

Figure 9. Average response times to detect the probe for each memory cue at each level of SOA for the Source Monitoring conditions. Error bars represent standard errors. R and F reflect data from the directed forgetting condition; A and B from the Arbitrary condition.

Multinomial modeling was conducted to augment the traditional data analysis techniques discussed above. Multinomial modeling provides a more sensitive analysis of memory performance by estimating of the contributions of individual cognitive processes to performance. Estimates of these processes are derived from Table 7, which includes the summed response frequencies for old and new items across DF and Arbitrary conditions. Before discussing the modeling results, it should be noted that the overall pattern of performance in Table 7 varies across the DF and Arbitrary conditions. This is informative for two reasons: first, it provides additional evidence that TSS is not an accurate account of intentional forgetting, and second, the pattern of errors in the DF condition is consistent with the pattern that would be expected if F items were truly forgotten. For example, items associated with an F cue at encoding are called "new" more often than they are called R. Additionally, when participants false alarm to a new item, they are more likely to call that an "F" item, relative to an "R" item. These results suggest that participants have difficulty discriminating between new and F items, as would be expected if these items were forgotten. These patterns were not present in the Arbitrary condition, where errors were similar across A and B responses.

Table 7

Response frequency matrices used for multinomial modeling

		Response						
		R	F	N		A	B	N
	R	846	358	263	A	656	425	263
Source	F	304	654	508	B	418	634	295
	N	195	630	2106	N	496	490	1710

Multinomial models generally include seven parameters that are believed to contribute to overall performance (Batchelder & Riefer, 1990). The parameters D_1 and D_2 reflect the ability to detect old items from new items, when those items are from Source_1 and Source_2, respectively. Parameters d_1 and d_2 are source discrimination estimates, and reflect the probability of labeling an item as belonging to Source_1 or Source_2 if the item was detected as old. In the current experiment, Source_1 and Source_2 estimates were calculated within memory cue conditions. Source_1 refers to R items and Source_2 refers to B items in the DF condition, while Source_1 and Source_2 refer to A and B items, respectively, in the Arbitrary condition. To illustrate, in the DF condition, the D_1 parameter reflects the ability to detect R items from new items, while d_1 indicates correct source identification for R items that were correctly called old. In addition to the mnemonic estimates, multinomial models often include three estimates of guessing**,** which were calculated per condition, not per source. The bias parameter, *b*, is defined as the probability of guessing "old" to a non-detected or new item. The other two parameters, *a* and *g*, reflect the probability that source discrimination is not known, but the participant guesses it belongs to Source_1: *a* is the parameter used when the item has been detected as "old", while *g* is the parameter used when the item was not detected as old. In many applications of multinomial modeling, *a* and *g* are set equal to one another, as guessing is typically not the psychological variable of greatest interest. Setting parameters to a constant value, or equal to one another, is often required, as a model with seven free parameters is often unidentifiable. In the current design, seven free parameters was untenable, as the number of free parameters outnumbered the six degrees of freedom that the data provided.

Batchelder and Riefer (1990) recommend conducting a set of tests to determine which parameter constraints are most warranted given a specific data set, and I followed their recommendations. First, the guessing parameters *a* and *g* were set to one another, as guessing was expected to be constant across detected and undetected items. Next, a chi-square test was conducted on the detection frequencies for R, F, A, and B items to determine if detection rates were equal across conditions. A significant difference was found, χ^2 (3, N = 332) = 126.01, p < .001, which indicates that D_1 and D_2 were not equal and should be allowed to vary in the models. The next step involved comparing discrimination rates across the various sources to determine which model fit better, a model that allowed the estimates to vary ($d_1 \neq d_2$) or a model that constrained the estimates to be equal ($d_1 = d_2$). These analyses required running both potential models and comparing the log likelihood ratios across the two models with a goodness of fit test (G^2). Small values of G^2 reflect small differences between predicted and observed probabilities, indicating a good fit, while significant values of G^2 indicate a poorer fit (Dodson, Prinzmetal, & Shimamura, 1998). The model that required $d_1 = d_2$ provided a poorer fit to the data, G^2 (1) = 106.76, p < .001, suggesting that these parameters should also be free to vary. Thus, the model used for final analysis had six free parameters: *D1*, *D2*, d_1, d_2, $a = g$, and *b*. Parameter estimates are presented in Table 8. Multinomial modeling analyses were conducted using both

Multitree (Moshagen, 2010) and Excel (Dodson et al., 1998) to confirm the accuracy of the modeling results.

Table 8

Estimated parameter values for multinomial model. Source 1 items were associated with the R cue in the DF condition and the A cue in the Arbitrary condition. Source 2 items were associated with the F cue in the DF condition and the B cue in the arbitrary condition.

	DF	Arbitrary
Parameter		
D_1	0.75	0.69
D_2	0.52	0.65
d_1	0.64	0.24
d_2	0	0.25
$a = g$	0.28	0.5
b	0.28	0.36

Note: DF = directed forgetting. D_1 = detection rate for source 1; D_2 = detection rate for source 2; d_1 = source identification rate for source 1; d_2 = source identification rate for source 2; a = parameter estimating the probability that a detected item with an unidentified source was attributed to source 1; g = parameter estimating the probability that an undetected item was attributed to source 1; b = parameter estimating the probability of calling an undetected old item or new item “old”.

These estimates revealed considerable deviations across the DF and Arbitrary conditions, but additional goodness of fit tests were conducted to determine if these differences were significant. The same model fitting procedure was used, wherein a given parameter estimate was set equal to the parameter estimate of interest, and goodness of fit (G^2) was compared. Given the large number of possible comparisons, only those comparisons that were of greatest theoretical importance were conducted. First, the probability of detecting an R item as old (D_1) was compared to the probability of detecting an F item as old

(D_2). The analysis revealed that these values were significantly different, G^2 (1) = 107.30, $p < .001$, consistent with previous DF effects. Next, the probability of detecting old A items was compared to detecting old B items, and no significant difference was found, G^2 (1) = 2.63, p = ns. Comparisons across source discrimination were also conducted. Discrimination for R (d_1) and F items (d_2) was significantly different, G^2 (1) = 105.8, $p < .001$, but the difference between A and B items was not, G^2 (1) = .01, p = ns. Thus, the multinomial modeling results replicate the typical recognition memory patterns, while indicating that source discrimination is adversely affected in the DF paradigm, contrary to what TSS predicts.

Discussion

Experiment 3 tested the TSS theory of intentional forgetting by comparing performance in DF conditions to conditions in which arbitrary labels were assigned and tagging was required. The rationale was that if TSS is the mechanism responsible for DF then data patterns should be similar across these conditions. For the sake of completeness, a variety of testing methods were included and dependent variables at encoding and retrieval were examined. First, the typical item-method procedure was compared across DF and Arbitrary conditions. The classic DF effect was observed in the DF condition, but no effect of memory cue was observed in the Arbitrary condition. These results were compared to a condition in which all words were to be remembered, but no memory cues were presented. If TSS were correct, performance should have been better in the Words Only condition, relative to the DF and Arbitrary conditions, as these conditions required twice as much information to be remembered. The corrected hit rates show that memory was worse in the Remember All condition when compared to memory for R items, but better when compared to memory for F items, replicating a pattern observed in a similar design (Muther, 1965). Furthermore, the corrected hit rates indicated that performance was better in Words Only condition, relative to the Arbitrary condition, suggesting that the tagging that occurred in the Arbitrary condition actually interfered with memory.

The data from the probe-detection task provided further evidence against TSS. The probe-detection task was embedded in the study phase of each condition so that estimates of effort during encoding could be compared. If tagging occurred for R and F items during encoding, then RTs to the probe task should have been similar across the DF and Arbitrary conditions. In fact, RTs were slower in the DF condition, relative to the Arbitrary and Words Only conditions. Additionally, the Memory Cue x Cue Condition interaction showed that RTs following F items were significantly slower than those following R items, while no difference was observed in the Arbitrary or Words Only conditions. These results suggest that the process that occurs during encoding varied across the three conditions: RTs were fastest in the Words Only condition, when participants could engage the same strategy on every trial. In the Arbitrary condition, participants were required to remember an additional piece of information, namely the memory cue, which resulted in slower responding to the probe. Finally, RTs in the DF condition were significantly slower than the Arbitrary condition, suggesting that something more than tagging was occurring. The pattern of RT results is consistent with the notion that DF instructions required participants to switch between remember and forgetting processes, and this switching results in significant slowing, compared to conditions in which forgetting is not required[7].

The data from the Inclusion conditions provide preliminary evidence that tagging is not a critical mechanism in DF. However, it has been argued that participants have reduced motivation to retrieve F items in standard recognition tests, and this could contribute to observed DF effects. To test this hypothesis, source monitoring and PDP testing conditions were included. These tests provided an opportunity for participants to use any tagged information they had stored about F items, as successful performance in both tasks required search for and use of all available information. If TSS were actually contributing to DF, these conditions should reveal evidence of this search process. Again, patterns differed significantly across the DF and Arbitrary conditions under both the source monitoring and PDP tests, providing additional evidence that TSS is not operating.

The primary purpose of Experiment 3 was to determine if TSS is a viable theory for DF. However, Experiment 3 also included two experimental techniques that had not yet been applied to DF. Inclusion and exclusion instructions from PDP were included at testing so that estimates of recollection and familiarity could be derived. As was done with the previous analyses, comparisons across DF and Arbitrary conditions were conducted, and the patterns varied across these conditions. Estimates of recollection were similar for F and B items, while

[7] This pattern was replicated in the Source Monitoring conditions, as well as in Experiment 2

estimates of familiarity were much higher for B items, relative to F items. The discrepancies across conditions seem to argue against TSS, but the recollection data are paradoxical. That is, the recollection estimates suggests that recollection contributed as much to memory for F items, which were to be forgotten, as it did for B items, which were to be remembered. This result appears to be consistent with TSS and suggests that some information about F items is stored.

Previous work comparing performance on PDP and source monitoring tasks suggests that the memory processes involved in PDP and source monitoring are the same (Buchner, Erdfelder, Steffens, & Martensen, 1997; Yu & Bellezza, 2000). Both tests require information about the source of an item for successful performance, but the contributions of these processes are estimated differently. PDP estimates of recollection are based on memory for the list on which an item appeared. However, a participant may recollect that an item was presented previously, but not be able to identify the list whence it came. Source monitoring models, like multinomial processing trees (MPTs), separate the contributions of recollection across item detection and source discrimination. MPTs go even further, and estimate the contributions of various guessing and bias terms to memory performance as well. Thus, source monitoring models provide a more sensitive description of how memorial processes influence recognition performance, and may augment the PDP data.

Parameter estimates in MPTs reflect the probability of making various responses, given that the higher-order response was made (Dodson et al. 1997). For example, in the current experiment the probability that source information contributed to accurate labeling of R items was .64, given that those items were previously identified as “old”. Consistent with the notion that intentional forgetting reduces memory for F items, the critical parameter estimates were lower for F items, relative to R items. First, the estimate of item detection (identifying an item as old or new) for R items (D_1) was higher than the estimate for F items (D_2), consistent with the typical DF effect (Table 8). Second, source discrimination was much higher for R items (d_1) relative to F items (d_2). In fact, the parameter estimate for d_2 was so small that it was essentially 0. This suggests that source memory does not contribute to the labeling of F items at all. This seems unlikely, as the data in the Table 7 clearly indicate that F items are often correctly identified, and the recollection estimate derived from PDP suggests that source memory does contribute to performance. To understand the implications of this value, it is helpful to consult the MPT for F items (Figure 5). The second tree reflects the possible ways that an F item could be classified, either correctly or incorrectly. The tree also indicates that an individual could arrive at a correct “F” decision in many ways. If the item is detected as old, the source could be remembered (d_2) or not remembered ($1 – d_2$). If it is not remembered, guessing can lead to an F response. On the other hand, if an item is not detected as old, a combination of guessing and bias can lead to an F response. The small value of the d_2 parameter indicates that source memory did not lead to the correct labeling of detected F items, but guessing and bias did. In other words, F judgments were based on factors other than explicit memory that the item was an F item. This is in stark contrast to the estimate of source discrimination for R items, which indicates that source memory played a considerable role in accurate classification of those items.

The multinomial modeling data suggests that memory for R items is quite good, while memory for F items is nearly eliminated. Guessing and bias contribute to both processes, but seem to have a larger influence on decisions about F items. The d_1 and d_2 parameter values for R and F items can also compared to the analogous estimates for A and B items. The recollection estimates obtained from the PDP conditions suggested that source memory contributed equally to F and B items. The MPT estimates indicate that source memory contributes more for B items than F items, while source memory for R items is better than for A items. To summarize, the PDP results appeared to offer support for TSS. However, a closer look at the contributions of biases and guessing via multinomial modeling indicated that source memory for F and B items was in fact quite different, again arguing against a tagging explanation.

It may seem surprising that factors like guessing and bias can account for so much of the source monitoring behavior, but evidence from other source monitoring tasks suggests that performance is dependent on metacognitive factors (Riefer, Hu, & Batchelder, 1994). Furthermore, source judgments can be influenced by manipulating stereotypes, metacognitive beliefs, or schemas (Broder & Meiser, 2007). The current evidence suggests that F items may be especially sensitive to these sorts of manipulations. The role of biases and decision-based factors has not been systematically examined in directed forgetting, but exploration of this area seems warranted for two reasons. First, the validity of the current results could be tested by including manipulations that should selectively affect the guessing and bias parameters, and determining if the effects influence the estimates as expected (Bayen, Murnane, & Erdfelder, 1996). Additionally, examination of these factors could provide insight into the mechanisms that contribute to apparent forgetting.

General Discussion

The current experiments were designed to test theories of directed forgetting in the item-method paradigm. Previous research suggested that forgetting in the item-method was due to selective rehearsal of R items, and passive decay of F items (Basden et al. 1993). This led other theories of item-method forgetting, such as attentional inhibition and tagging and selective search (TSS), to be largely ignored for the past decade. Recent evidence examining the effort required to remember and forget argues against a passive decay account, and has renewed interest in theories that view forgetting as an active, effortful process (Fawcett & Taylor, 2008). Two such theories were scrutinized in the current set of experiments: attentional inhibition and TSS. Across three experiments, a variety of stimuli, encoding manipulations, and testing methods were included to test each perspective. In Experiment 1, manipulations of level of processing and cue timing were included in an attempt to distinguish between competing predictions made by the two theories. Unfortunately, the results were generally inconclusive, as the required interactions did not appear. Experiments 2 and 3 examined the critical aspects of attentional inhibition and TSS separately. In Experiment 2, a probe-detection task was embedded into the DF task and three classes of stimuli (words, images, faces) were compared. RTs across R and F faces offered a means of comparing two versions of attentional inhibition theory, representation suppression and rehearsal suppression. Probe RTs were longer to F faces, relative to F faces, consistent with representation suppression. Finally, Experiment 3 tested the tenets of TSS by comparing performance in DF conditions to performance in conditions where participants were encouraged to adopt a TSS approach. Dependent measures at encoding and testing were examined, and results revealed disparate patterns of performance across DF and Arbitrary conditions, discounting the role of TSS in DF.

What do the combined results tell us about the nature of forgetting? First, DF effects were consistently observed, supporting the notion that intentional forgetting is a robust phenomenon that appears across a variety of stimuli and testing procedures. Concern about the role demand characteristics play in DF effects seems unwarranted, given the variety of experimental situations in which they were observed. For example, testing procedures capable of providing more sensitive measures of forgetting, such as PDP and source monitoring, were included and significant DF effects emerged in these conditions as well as standard recognition tests. Even if a participant was able to guess the hypothesis of the DF experiments, it is unlikely they were able to tailor their performance under such complex testing instructions.

With respect to theories of DF, the results are suggestive. Although Experiment 1 was inconclusive, the body of evidence from Experiment 3 argues against a TSS account of intentional forgetting. In Experiment 3, DF conditions were compared to conditions in which participants had to remember arbitrary memory cues. Successful performance in the Arbitrary conditions thus depended on the ability to tag words with the appropriate memory cue. If TSS were contributing to DF performance, patterns should have been similar across the DF and Arbitrary conditions. Data from the probe detection task that was included during encoding revealed that the processing during DF varied from the processing that occurred when arbitrary labels were tagged to words. Recognition data from the inclusion, source monitoring, and PDP tasks indicated unique processing during testing as well. Furthermore, the results from multinomial modeling indicated that forgetting impacted memory in a striking way: guessing and bias had more of an influence on source memory than memory for the tag did. The evidence from Experiment 3 is consistent and compelling, and suggests that TSS is not a viable explanation for intentional forgetting.

Additionally, the results confirm that forgetting is an effortful process. Experiments 2 and 3 included probe detection tasks that were used to estimate the effort required to remember and forget across various manipulations. Although absolute RTs differed across stimulus classes and encoding instructions, the patterns for DF trials were consistent across both experiments: RTs to F items were significantly slower than RTs to R items. This suggests that forgetting is not due to a passive decay process, as the selective rehearsal theory states (Bjork, 1972). The results are consistent with an attentional inhibition theory, which argues that a demanding process acts to reduce accessibility of F items in memory. In Experiment 2, two variations of the attentional inhibition theory were compared. Rehearsal suppression states that attentional resources act to suspend the rehearsal process for F items (Hasher et al. 1996), while representation suppression states that an active attentional process acts to suppress the representation of F items in memory (Fawcett & Taylor, 2008). To compare these two explanations, pictures of non-famous faces, which are less amenable to verbal rehearsal, were included and probe RTs were compared. Representation suppression predicted that RTs to F faces would be slower, relative to R faces, while rehearsal suppression predicted no difference. The data revealed that RTs following F faces were significantly slower, relative to R faces, which argues for representation suppression.

The supposition that rehearsal suppression is not the critical mechanism in DF depends on accepting the assumption that faces are not selectively rehearsed. This is a strong claim, and may require further examination. Several methods of additional testing could be used to evaluate such a statement. First, “think aloud” experiments, wherein participants are asked to verbalize their thinking during the encoding phase of a DF experiment, would provide preliminary evidence for the presence of verbal (or non-verbal) rehearsal for faces, images and words. A second option is to utilize the same procedure described in Experiment 2, but include pictures of famous faces. Famous faces share all the same complexities as non-famous faces, but they also are associated with a label (the name) that makes verbal rehearsal possible. If rehearsal contributes to probe-detection and recognition performance, then the evidence from the famous faces condition should match the patterns observed in the words and images conditions.

Specifically, RTs should be longer and memory should be better in the famous faces condition, relative to the non-famous face condition. Although not definitive, this would provide additional evidence that rehearsal contributes differently for faces relative to other stimuli.

As mentioned, the probe RT data argue against a passive decay of forgetting, the explanation provided by selective rehearsal theory. Two additional findings further challenge the selective rehearsal account. First, if rehearsal is a critical component of DF, then manipulations that occur at encoding should influence the size of the DF effect. Level of processing is a robust encoding manipulation in memory research, and was included in Experiment 1 (Craik & Lockhart, 1972). If memory for R items is dependent on rehearsal at encoding, then deep processing should have improved memory for R items, relative to shallow processing. However, memory for R items was equal across deep and shallow conditions. Evidence from the faces condition of Experiment 2 provides a second challenge to selective rehearsal. It was argued that faces are not rehearsed, and the data from the probe RT and recognition tests were consistent with this notion. The DF effect still emerged, with memory for R faces surpassing memory for F faces. It is unclear how selective rehearsal theory would explain the DF effect in faces. That is, selective rehearsal theory states that the DF effect results from the rehearsal benefit that R items receive. If rehearsal is not occurring, then what mechanism would lead to improved memory for R faces, relative to F faces? Representation suppression posits an active suppression process, which would reduce memory for F faces, regardless of the processing that occurred for R faces.

Therefore, the DF effect is plausible given a representation suppression perspective, but difficult to reconcile with selective rehearsal.

Although these results are inconsistent with selective rehearsal theory, a complete dismissal of the theory may be unwarranted. Selective rehearsal has dominated the item-method DF literature for several reasons. First, as mentioned in the introduction, evidence from several prominent experiments supports this theory (Basden et al.1993; Wetzel & Hunt, 1977). Second, it is intuitive. Memory researchers have long promoted the important role that rehearsal, especially elaborative rehearsal, plays in memory (Craik & Tulving, 1975; Rundus, 1971). Furthermore, evidence from "think-aloud" experiments confirms that participants do differentially rehearse R and F items (Golding, Roper, & Hauselt, 1996). Given these points, it may be imprudent to entirely discount the role that selective rehearsal of R items plays in DF, despite the current inconsistencies. However, the evidence clearly suggests that the passive decay explanation should be seriously reconsidered.

The current probe RT data add to a growing set of evidence suggesting that attentional resources are required to intentionally forget. Additional behavioral evidence has come from DF experiments examining inhibition of return (Taylor, 2005), go/no-go tasks (Hourihan & Taylor, 2006), and the influences on conscious and unconscious memory (David & Brown, 2003). Physiological evidence has also been found. For example, Wylie, Foxe and Taylor (2008) used fMRI to compare patterns of activation when forgetting was intentional and unintentional. They used an item-method design to present R and

F cues at encoding, and then used recognition performance to analyze activity at encoding, post-hoc. Intentional and unintentional forgetting was examined by comparing correct rejections for F items with misses to R items. Unintentional and intentional remembering were examined by comparing hits to F and R items, respectively. BOLD signal changes during trials that led to these responses were then analyzed: Intentional forgetting trials were associated with unique neural substrates when compared with intentional remembering and unintentional forgetting trials. Specifically, intentional forgetting was associated with increased activity in frontal areas that are associated with cognitive control, like cingulate cortex, relative to unintentional forgetting. Furthermore, intentional forgetting was associated with unique areas in medial temporal lobe, relative to intentional remembering. These results do not provide clear evidence that cognitive control mechanisms are engaged when an intention to forget is active. However, the presence of additional activity in frontal and temporal lobes argues against a passive decay explanation, and for an active process that may act in tandem with remembering processes.

The current results suggest that representations of F items are suppressed in memory. However, the nature of this suppression process is unclear. Suppression and inhibition are concepts that have been espoused in theories of memory and cognitive control, but are often poorly defined (Anderson, Bjork, & Bjork, 1994; Burgess & Shallice, 1996; Kane, Bleckley, Conway & Engle, 2001). Most theories of suppression in memory invoke a top-down, executive control mechanism that acts to reduce accessibility of particular items (Anderson &

Green, 2003). Within the DF literature, several explanations of suppression have been provided. For example, Zacks et al. (1996) suggested that inhibition acts to remove an item from memory and prevent any future intrusions. Quinlan et al. (2010) stated that inhibition prevents commitment of items to long-term memory. It may also be the case that executive control processing reduces activation of F items so that they are less accessible in memory later on (Anderson, Bjork, & Bjork, 1994). While all are viable explanations, these hypotheses do little to advance our understanding of precisely how suppression impacts performance.

Gieselman and Bagheri (1985) provided a more concrete hypothesis. In an attempt to explain the role retrieval inhibition plays in DF, they suggested that "a forget command initiates a process that serves to inhibit or block access routes to target items in episodic memory" (p 62). This process would prevent retrieval of items at testing, as the paths would be degraded. The idea of inhibition at retrieval may seem inconsistent with the probe-detection evidence observed in Experiments 2 and 3, which implicates processing at encoding as a contributing factor to intentional forgetting. Despite the name, Gieselman and Bagheri's retrieval inhibition explanation posits a suppression process that occurs at encoding. They suggested that an F cue initiates an inhibitory process, and acts to prevent retrieval later on. If true, this process must occur at encoding, when F cues are provided. This explanation wasn't directly tested in the current experiments, but it appears to be similar (if not entirely consistent) with the representation suppression theory that Fawcett and Taylor (2008) articulated. They suggested that representations of F items are suppressed at encoding via an effortful inhibitory process, which leads F items to be underreported at testing. A process that acts to break the connections between memories would be effortful, and provides one potential mechanism for a suppression explanation. Furthermore, if the retrieval paths for F items are disrupted, this may make the intact paths to R items more accessible, which could explain the benefit that R items receive in DF (Experiment 3). This explanation provides one potential mechanism for suppression in DF, but additional testing is necessary to fully evaluate this hypothesis.

Additional testing is also necessary to understand the role that metacognitive factors play in DF experiments. When participants were asked to provide the source of old items in Experiment 3, F items were often correctly identified, but the estimate of source discrimination (d_2) derived from multinomial modeling suggested this was not due to retained memory for the source. Other factors like bias and guessing led participants to identify F items correctly, despite a lack of explicit source memory. The negligible value of d_2 was surprising, and additional experimentation is necessary. Replication of the current experiment would be helpful to ensure that the current parameter estimates were not the result an extreme sample. However, more compelling evidence would come from manipulations of bias and guessing (Bayen et al., 1996). If bias and guessing have an impact on memory for F items, then manipulating these factors systematically should lead to changes in source detection. Of course, bias and guessing processes also contributed to memory for R items, but the considerable estimate of discrimination for R items (d_1)

suggested that source memory has a larger influence on the accurate labeling of R items than bias and guessing do. Thus, the current results would lead to the following prediction: Manipulations of guessing and bias should have a larger impact on source discrimination for F items, but may not have an effect on R items. This is an area that has not been systematically explored in DF literature, and provides an opportunity for future research.

In conclusion, the current evidence indicates that forgetting in the item-method design is not the result of differential rehearsal, and may be due to an active inhibitory process. Recently, the list-method of DF has received the majority of the attention in the DF literature, because forgetting in the item-method was believed to be due to an "uninteresting" decay process (Epstein, 1972 pg 151; see also Johnson, 1994). For example, Sahakyan and Delaney (2005) stated that the mechanism responsible in the list-method of DF "has created a variety of theoretical viewpoints as opposed to the mechanism supporting the item method" (p. 789). Their comment illustrates how alternatives to selective rehearsal, like attentional inhibition, have been neglected in DF. Perhaps this is because the original conceptualization of attentional inhibition, rehearsal suppression, also implied a passive decay process when taken to its logical extreme. Zacks et al. (1996) stated that the rehearsal process was suppressed for F items, which suggests that the lack of rehearsal eventually leads to decay from memory. According to this account, rehearsal suppression is effortful because a secondary inhibitory process needed to be engaged to stop an ongoing process, not because the process of forgetting is active. The representation suppression theory makes the opposite case: forgetting is a conscious, effortful act that requires resources to attempt to suppress items in memory. The current results suggest that forgetting is effortful, which may align the item-method with the list-method more than originally thought (Basden et al., 1993). Although multiple theories of forgetting in the list method also exist, many invoke an inhibitory mechanism that contributes to forgetting via disrupted retrieval. The current evidence suggests a similar inhibitory mechanism may function during encoding in the item-method.

References

Baddeley, A. D. (1986). *Working Memory*. New York: Oxford University Press.

Basden & Basden (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory, 4*, 633 – 653.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology:Learning, Memory and Cognition, 19,* 603 – 616.

Batchelder, W. H. & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548 – 564.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection and multinomial modeling of source monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 197 – 215.

Bjork, R. A. (1970). Positive forgetting: The non-interference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9*, 255 – 268.

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217 – 235). Washington DC: Winston.

Bjork, R. A., LaBerge, D. & Legrand, R. (1968). The modification of short term memory through instructions to forget. *Psychonomic Science, 10,* 55 – 56.

Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, U. S. A.*, 105 (38), 14325 – 14329; doi: 10.1073/pnas.0803390105

Broder, A. & Meiser, T. (2007). Measuring source memory. *Journal of Psychology, 215*, 52 – 60. doi: 10.1027/0044-3409.215.1.52

Buchner, A., Erdfelder, E., Steffens, M. C., & Martensen, H. (1997). The nature of memory processes underlying recognition judgments in the process dissociation procedure. *Memory and Cognition, 25*, 508 – 517.

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.

Craik, F. I. M., & Tulving (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268 – 294.

Dodson, C. S., Prinzmetal, W., & Shimamura, A. P. (1998). Using Excel to estimate parameters from observed data: An example from source memory data. *Behavior Research Methods Instruments and Computers, 30*, 517 – 526.

Ellis, H. D. (1975). Recognising faces. *British Journal of Psychology*, *66*, 409 – 426.

Elmes, D. G., Adams, C. A., III & Roediger, H. L. III (1970). Cued forgetting in short-term memory: Response selection. *Journal of Experimental Psychology, 86*, 103 – 107.

Ekman, P., & Matsumoto, D. (1993). Japanese and Caucasian Neutral Faces (JACNeuF). Photographs on CD-Rom

Epstein, W., Massaro, D. W., & Wilder, L. (1972). Selective search in directed forgetting. *Journal of Experimental Psychology*, *94*, 18-24.

Epstein, W., & Wilder, L. (1972). Searching for to-be-forgotten material in a directed forgetting task. *Journal of Experimental Psychology, 95, 349 -357.*

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. doi: 10.3758/MC.36.6.1168

Geiselman, R. E., (1975). Semantic positive forgetting: Another cocktail party phenomenon. *Journal of Verbal Learning and Verbal Behavior, 14*, 73 – 81.

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition, 13*, 57-62

Goernert, P. N., Widner, R. L., & Otani, H. (2007). Classification accuracy across multiple tests following item method directed forgetting. *Quarterly Journal of Experimental Psychology*, *60*, 1178 – 1186. doi: 10.1080/17470210701396848

Hasher, L. & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation*. (Vol 22, pp 193 – 223). San Diego: Academic Press.

Hauswald, A. & Kissler, J. (2008). Directed forgetting of complex pictures in an item method paradigm. *Memory*, *16*, 797 – 809. doi: 0.1080/09658210802169087

Homa, D., & Spieker, S. (1974). Assessment of selective search as an explanation for intentional forgetting. *Journal of Experimental Psychology, 103*, 10 – 15.

Horton, N., Hay, D. C., & Smyth, M. M. (2008). Hebb repetition effects in visual memory: The roles of verbal rehearsal and distinctiveness. *The Quarterly Journal of Experimental Psychology, 61*, 1769 – 1777. doi:10.1080/17470210802168674

Horton, K. D., & Petruk, R. (1980). Set differentiation and depth of processing in the directed forgetting paradigm. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 599 – 610.

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

Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual-learning. *Journal of Experimental Psychology: General*, *110*, 306 – 340.

Johnston, C. D. & Jenkins, J. J. (1971). Two more incidental tasks that differentially affect associative clustering in recall. *Journal of Experimental Psychology, 89*, 92 – 95.

Johnson, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin, 16*, 274 – 292.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1994). Source monitoring. *Psychological Bulletin*, *114*, 3 – 28.

Kausler, D. H. & Settle, A. V. (1975). The false-recognition effect for homophones of items subjects are cued to forget. *American Journal of Psychology, 88*, 627 – 634.

Keppel, G. & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook.* New Jersey: Pearson Prentice Hall.

Lee, Y. & Lee, H. Divided attention facilitates intentional forgetting: Evidence from item-method directed forgetting. *Consciousness and Cognition*, (2010), doi:10.1016/j.concog.2010.09.008

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, *104*, 271 – 279.

MacLeod, C. M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 13 – 21.

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 and Review, 6*, 123 – 129.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, and Computers, 36,* 630 – 633.

Moshan, M (2010). multiTree: A computer program for the analysis of multnomial processing tree models. *Behavior Research Methods, 42*, 42 – 54. doi: 10.3758/BRM.42.1.42

Muther, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science, 3*, 429 – 430.

Paller K. A., Bozic, V. S., Ranganath, C., Grabowecky, M., Yamada, S. (1999). Brain waves following remembered faces index conscious recollection. *Cognitive Brain Research, 7*, 519 – 531.

Pashler, H. (1989). Dissociations and dependencies between speed and accuracy – Evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology, 21*, 469 – 514.

Pashler, H. (1994). Dual-task interference in simple tasks – Data and theory. *Psychological Bulletin, 116*, 220 – 244.

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face recognition algorithms. *IEEE Trans Pattern Analysis and Machine Intelligence, 22,* 1090 – 1104.

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. (1998). The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing Journal, 16*, 295 – 306.

Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce more rapid progress than others. *Science, 146*, 347 – 353.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychological Software Tools.

Quinlan, C. K., Taylor, T. L., & Fawcett, J. M. (2010). Directed forgetting: Comparing pictures and words. *Canadian Journal of Experimental Psychology*, *64*, 41 – 46. doi 10.1037/a0016569

Reitman, W., Malin, J. T., Bjork, R. A., & Higman, B. (1973). Strategy control and directed forgetting. *Journal of Verbal Learning and Verbal Behavior, 12,* 140 – 149.

Riefer, D. M., Hu, X. & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 680 – 693.

Roediger, H. L., III (1990). Implicit memory: Retention without remembering. *American Psychologist, 45*, 1043 – 1056.

Sahakyan, L. & Delaney, P. F. (2005). Directed forgetting incidental learning and recognition testing: Support for a two-factor account. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 789 – 801. doi: 10.1037/0278-7393.31.4.789

Shepard, R.N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior, 80,* 1 – 24.

Shebilske, W. Wilder, L. & Epstein, W. (1971). Forget instructions: Effect of selective rehearsal and categorical distinctiveness. *Journal of Experimental Psychology, 87*, 372 – 378.

Smyth, M. M., Hay, D. C., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual spatial domain: Reconstructing sequences of unfamiliar faces. *Quarterly Journal of Experimental Psychology, 58A*, 909. doi: 10.1080/02724980443000412

Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207 – 222.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652 – 654.

Wetzel, C. D. (1975). Effect of orienting tasks and cue timing on the free recall of remember- and forget-cued words. *Journal of Experimental Psychology: Human Learning and Memory*, *1,* 556 – 566.

Wetzel, C. D & Hunt, R. E. (1977). Cue delay and the role of rehearsal in directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, *3,* 233 – 245.

Wilson, M.D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments, and Computers*, *20*, 6-11.

Winer, B. J. (1971) *Statistical Principles in Experimental Design*, 2nd ed. New York: McGraw-Hill.

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

Yu, J. & Bellezza, F. S. (2000). Process dissociation as source monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1518 – 1533. doi: 10.1037//0278-7393.26.6.1518

Zacks, R. T., Radvansky, G., and Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 143 – 156.

APPENDIX A

STIMULI USED IN EXPERIMENT 1.

ANT
ANTELOPE
APARTMENT
APE
APRON
ARMADILLO
ARTICLE
BABOON
BAG
BALL
BANDAID
BEAR
BEAVER
BED
BEE
BEER
BEETLE
BELT
BICYCLE
BINDER
BLOUSE
BOBCAT
BOOK
BOOT
BOTTLE
BRACELET
BROOM
BUFFALO
BUS
BUTTER
BUTTERFLY
BUTTON
BUZZARD
CAMEL
CAMERA
CAN
CANNISTER
CAP
CAR
CARD
CARDINAL
CARIBOU
CARPET
CAT
CEILING
CENTIPEDE
CHAIR
CHEETAH
CHICKEN
CHIMPANZEE
CHIPMUNK
CHIPS
CLIP
CLOCK
COAT
COFFEE
COIN
CONDOR
COOLER
CORD
COUCH
COW
COYOTE
CRAB
CRICKET
CROCODILE
CROW
CROWN
CUP
DEER
DESK
DOG
DOLPHIN
DOOR
DOVE
DRESS
DUCK
EAGLE
EEL
ELEPHANT
ENVELOPE
FALCON
FAN
FERRET
FLAMINGO
FLOOR
FOOTBALL
FORK
FOX
FRAME
FROG
GAME
GASOLINE
GAZELLE
GERBIL
GIN
GIRAFFE
GLASS
GLOVE
GOAT
GOOSE
GOPHER
GORILLA
GRASSHOPPER
HALIBUT
HAMSTER
HAT
HAWK
HIPPO
HORSE
HOUSE
HUMAN
HYENA
IGUANA
JACKET
JAGUAR
JERSEY
JUICE
KANGAROO
KEY
KEYBOARD
KITTEN
KNIFE
LAMB
LAMP
LEMONADE
LEMUR
LEOPARD
LIGHT
LION
LIQUOR
LIZARD
LLAMA
LOBSTER
LOCK
MAGAZINE
MAGNET
MANATEE
MARKER
MEDAL
MILK
MONKEY
MOOSE
MOSQUITO
MOTH
MOUSE
MUG
NAIL
NOTE
OCTOPUS
OSTRICH
OTTER
OUTLET
OWL
PAINTING
PANDA
PANTS
PAPER
PARAKEET
SCOOTER
SCORPION
SCREW
TOASTER
TRAIN
TRASH

PARROT
PATIO
PELICAN
PEN
PENCIL
PHEASANT
PHONE
PHOTO
PIG
PIGEON
PILL
PIN
PLATE
PONY
PORCUPINE
POSTER
POT
PRINTER
PUMA
PURSE
QUAIL
RABBIT
RACOON
RAG
RAT
RHINOCEROUS
RING
ROADRUNNER
ROBE
ROBIN
RUG
RUM
SALMON
SAND
SANDAL
SCISSORS
SEAGULL
SEAL
SHAMPOO
SHARK
SHEEP
SHELF
SHIRT
SHOE
SHRIMP
SIGN
SKATES
SKIRT
SKUNK
SNAKE
SOAP
SOCK
SODA
SPARROW
SPEAKER
SPIDER
SPOON
SQUID
SQUIRREL
STAPLER
STINGRAY
STORK
STREET
SUIT
SWAN
TABLE
TAPE
TEA
TERMITE
TIE
TIGER
TOAD
TROUT
TRUCK
TUNA
TURKEY
TURTLE
VIPER
VULTURE
WASP
WATER
WEASEL
WHALE
WHISKEY
WINE
WOLF
WOMAN
WOODPECKER
WORM
ZEBRA
ZIPPER

APPENDIX B

STIMULI USED IN THE WORD CONDITION OF EXPERIMENT 2.

BAGEL	*GLASS*	*OVEN*
BALL	*GUITAR*	*PEPPERS*
BANANA	*HANGER*	*PHONE*
BARREL	*HAT*	*PIZZA*
BASKET	*HORN*	*RAKE*
BEAKER	*KEYBOARD*	*RIBBON*
BOX	*LADDER*	*ROSE*
BREAD	*LADDLE*	*RULER*
BULLET	*LANTERN*	*SCALE*
CAKE	*LEAF*	*SCARF*
CALCULATOR	*LIGHT*	*SHIP*
CANTALOUPE	*LIPSTICK*	*STOOL*
CHESS	*LOBSTER*	*SUITCASE*
CHISEL	*LOCK*	*SUNGLASSES*
DICE	*MICROPHONE*	*SWEATER*
DINOSAUR	*MIRROR*	*TIGER*
DRESSER	*MONEY*	*TRAIN*
EGG	*NAPKIN*	*TROPHY*
FEATHER	*NEEDLE*	*TURTLE*
FOLDER	*NOTEBOOK*	*UMBRELLA*

APPENDIX C

STIMULI USED IN EXPERIMENT 3.

ALCOHOL
ANIMAL
BAG
BALL
BANANA
BATH
BED
BIN
BIRD
BLUE
BOX
BREAD
CHAIR
CIRCLE
CITY
CLOCK
CLOTHING
COAT
COST
COUNTRY
COVER
DATE
DESK
DINNER
DISEASE
DOOR
DRINK
DRIVER
EGG
END
ESSAY
EXIT
EYE
FACT
FAMILY
FILM
FINISH
FIRE
FUTURE
GLASS
HALL
HAND
HAT
HEAD
HEAT
HILL
HOLD
HOME
HOUR
HOUSE
HUNGER
JACKET
JOB
JOKE
KEY
KISS
KNEE
LAKE
LAMP
LEG
LETTER
LIBRARY
LIKE
LOCK
MAGAZINE
MALE
MIRROR
MISTAKE
MONEY
NIGHT
NOTE
NUMBER
PAGE
PAINTER
PENCIL
PEOPLE
PERSON
PIE
PILLOW
POST
POT
POTATO
POUND
PROBLEM
RADIO
RAIN
RECORD
REST
SALT
SCHOOL
SEAT
SHAPE
SHIRT
SINK
SISTER
SKIN
SKY
SLEEP
SMART
SMILE
SOAP
SONG
SPOON
SPRING
STAFF
STREET
TALK
TEETH
TICKET
TOE
TOWN
TREE
TURN
VOICE
WALL
WANT
WATCH
WEEK
WINDOW
WORLD

ARIZONA STATE UNIVERSITY
RESEARCH AND ECONOMIC AFFAIRS

Office of Research Integrity and Assurance

To:	Stephen Goldinger PSY
From:	Mark Roosa, Chair Soc Beh IRB
Date:	01/15/2010
Committee Action:	**Expedited Approval**
Approval Date:	01/15/2010
Review Type:	Expedited F7
IRB Protocol #:	1001004703
Study Title:	Directed Forgetting
Expiration Date:	01/14/2011

The above-referenced protocol was approved following expedited review by the Institutional Review Board.

It is the Principal Investigator's responsibility to obtain review and continued approval before the expiration date. You may not continue any research activity beyond the expiration date without approval by the Institutional Review Board.

Adverse Reactions: If any untoward incidents or severe reactions should develop as a result of this study, you are required to notify the Soc Beh IRB immediately. If necessary a member of the IRB will be assigned to look into the matter. If the problem is serious, approval may be withdrawn pending IRB review.

Amendments: If you wish to change any aspect of this study, such as the procedures, the consent forms, or the investigators, please communicate your requested changes to the Soc Beh IRB. The new procedure is not to be initiated until the IRB approval has been given.

Please retain a copy of this letter with your approved protocol.

COVER LETTER
Directed Forgetting

Date: January 11, 2010

Dear Participant:

I am a graduate student under the direction of Professor Goldinger in the Department of Psychology at Arizona State University.

I am conducting an experiment to study memory. I am inviting your participation, which will involve remembering and forgetting information that will appear on the screen. You will be given a memory test later on, to determine how well you were able to remember and forget.

Your participation in this study is voluntary. If you choose not to participate or to withdraw from the study at any time, there will be no penalty, (for example, it will not affect your grade). You must be 18 or older to participate in the study.

Although there is no immediate benefit to you, possible benefits of your participation include a contribution to knowledge of cognitive processes that may impact future research or theory. There are no foreseeable risks or discomforts to your participation. You will be awarded credit for your participation in this experiment.

Your responses will be confidential. The data you provide will not be attached to your name, and will be kept safely in locked facilities. The results of this study may be used in reports, presentations, or publications but your name will not be known. Results will only be shared in the aggregate form.

If you have any questions concerning the research study, please contact the research team at: either Goldinger@asu.edu or Whansen@asu.edu. If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480) 965-6788.

Return of the questionnaire will be considered your consent to participate.

Sincerely,

Whitney Hansen

ASU IRB
Approved
Sign Sm
Date 11510 - 11411