On the Cost and Benefit of Taking it out of Context: 
Modeling the Inhibition Associated with Directed Forgetting

Kenneth J. Malmberg (malmberg@cas.usf.edu)
University of South Florida
Department of Psychology, 4202 East Fowler Ave., PCD 4118G
Tampa, FL 33620 USA

Melissa Lehman (mlehman@mail.usf.edu)
University of South Florida
Department of Psychology, 4202 East Fowler Ave., PCD 4118G
Tampa, FL 33620 USA

Lili Sahakyan (l_sahaky@uncg.edu)
University of North Carolina
Department of Psychology
296 Eberhart Building, PO Box 26170
Greensboro, NC 27402 USA

Abstract

Forgetting can occur as the result of unconscious or automatic memory processes or as the result of conscious control. The later form of forgetting is often referred to as suppression, repression, or inhibition, and it is investigated in the laboratory using the directed forgetting procedure. The authors describe and empirically test the first formal model of directed forgetting, implemented within the framework of the Search of Association Memory Theory (SAM). The critical assumption is that episodic memory can be suppressed by a conscious attempt to alter the mental context in which new memories are encoded. The present model accounts for both veridical and erroneous free recall performance.

Keywords. Memory; Cognitive Models

Directed Forgetting

Any spouse, partner, graduate student, or even Sigmund Freud would tell you that the notion that people intentionally forget prior events is not new. While Freud posited that the forgetting of negative experiences is a mechanism for alleviating stress or anxiety, some memory theorists have posited that forgetting some experiences is necessary in order to remember other experiences (e.g., Bjork, LeBerge, & Legrand, 1968). How, not why, one intentionally forgets is the subject of this paper.

Intentional forgetting is investigated in the laboratory using two directed forgetting methods: the item and the list methods. In the item method, subjects study a list of items, and for each item they are informed if their memory for that item will be later tested. Importantly, memory is tested for all items, regardless of whether they were told during study that the items were to be remembered or not. For instance, memory might be tested using a free recall procedure, whereby subjects are asked to generate as many items from the list as they can in any order. The common finding is that the to-be-remembered items are better recalled than the to-be-forgotten items.

Here we utilize the list method of directed forgetting, although we will discuss the item method in the General Discussion. Traditionally, subjects study two lists of items ($L_1$ and $L_2$). After studying $L_1$, subjects are sometimes told that their memory for $L_1$ will not be tested. Much to the chagrin of those in the forget condition, subjects are then told at test to recall items from both lists. There are two effects of the “forget” instruction when memory for both lists is subsequently tested via free recall: $L_1$ items are remembered worse and $L_2$ items are remembered better in the forget $L_1$ condition. That is, $P(L_1 \text{ recall} – \text{ forget}) < P(L_1 \text{ recall} – \text{ remember})$ and $P(L_2 \text{ recall} – \text{ forget}) > P(L_2 \text{ recall} – \text{ remember})$. These effects are referred to as the cost and the benefit of directed forgetting, respectively (Sahakyan & Kelley, 2002).

Modeling Directed Forgetting

There are several traditional explanations of directed forgetting (MacLeod, 1998 for a review). $L_1$ items might be rehearsed less often following forget instructions. Alternatively, some suggest that the cost and benefit of directed forgetting is due to output interference; $L_2$ items tend to be recalled before $L_1$ items, and this creates additional interference when subjects try to recall $L_1$ items (Geiselman, Bjork, & Fishman, 1983). Presumably, this tendency increases as a result of forget instructions.

While differential rehearsal of to-be-remembered and to-be-forgotten items might occur for the item method, it is unlikely to provide a complete explanation of directed forgetting for the list method for reasons that will be made clear momentarily. In addition, it is unclear why $L_2$ items would tend to be output first to a greater degree after forget instructions than after remember instructions.
A different explanation of directed forgetting is that \( L_1 \) items are suppressed or inhibited as a result of the forget instruction, but this is only a little more than a description of the data. The model that we describe can be considered a model of the mechanisms involved in suppression or inhibition phenomena. It is a variant of the set-differentiation (SD; Bjork, 1972) originally described by Sahakyan and Kelley (2002), which assumes that the forget instruction causes items to be encoded in memory in a manner in which \( L_1 \) and \( L_2 \) can be more easily differentiated during memory testing. A forget instruction reduces the similarity between the stored \( L_1 \) and \( L_2 \) context information, information that allows for episodic memory of items encountered thousands of times in everyday life. We will refer to this as context differentiation (or CD)

This assumption was tested by Sahakyan and Kelley (2002; also Sahakyan, 2004). In these experiments, some subjects participated in a traditional directed forgetting experiment, while other subjects were induced to “change contexts” between \( L_1 \) and \( L_2 \). Prior to \( L_1 \) subjects might have been told to, “Imagine that you are on the moon”, and after \( L_1 \) subjects were either given the same imagery instruction or they were told to, “Imagine that you are invisible”. Sahakyan and Kelley assumed that the change in mental imagery instructions would decrease the similarity between the \( L_1 \) and \( L_2 \) contexts. If so, and if the SSDS assumption is correct, then the cost and benefit associated with forget instructions should also be observed with instructions to change contexts. Their results supported this version of the assumption and challenged the rehearsal accounts of directed forgetting.

**SAM Model of Free Recall**

To provide a concrete explanation of directed forgetting, we work within the framework of the Search of Associative Memory theory (SAM; Raaijmakers & Shiffrin, 1980). According to SAM, episodic traces (i.e., images) consist of a set of associations between images and contexts, with the strength of the associations dependent on how long items were studied and what items were rehearsed together either recently or in the past. The parameter, \( a \), is the strength of association between an item and a contextual cue, referred to as context strength. The parameter, \( b \), is the strength of association between two items that were recently rehearsed together when one item is used to cue the image of the other item (inter-item strength). The parameter, \( c \), is referred to as self strength, and it is the associative strength between an item’s image and the same item used as a cue. Lastly, the parameter, \( d \), is the strength of association between two items that were not recently rehearsed together, and it is referred to as residual strength.

The context strength parameter, \( a \), is the key parameter for implementing the CD assumption (Ratcliff, VanZandt, & McKoon, 1995; Shiffrin, Ratcliff, & Clark, 1990; Malmberg & Shiffrin, 2005). In fact, the set of CD assumptions that we will describe make no reference to \( b, c, \) or \( d \), and hence they can be ignored. A more complex model would require assumptions about rehearsal and the use of different types of retrieval strategies, and we discuss several possibilities in the General Discussion when we consider how the present model might be extended to account for the item method of directed forgetting. For now, it is sufficient to note that the SAM model of free recall assumes that retrieval consists of sampling images from memory and attempting to recover the information that they contain. The probability of sampling image, \( I \), given \( Q \) as a retrieval cue is:

\[
P(I,Q) = \frac{S(I,Q)}{\sum_{j=1}^{m} S(J,Q)}
\]

where \( m \) images are stored, and \( S(J,Q) \) is the strength of association between the retrieval cue and image, \( J = 1...m \).

We assume that each probe of memory is with a context cue only. While SAM assumes that both item and context cues can be used to probe memory, we make the simplifying assumption that \( b \) is the same for all images, and hence item cues do not differentially affect directed forgetting. Thus, the present model attempts to account for directed forgetting using the list method without appeal to a rehearsal account.

Since the list method involves studying more than one list, we assume that the context changes between them (cf. Mensink & Raaijmakers, 1989). Call the lists \( L_x \) and \( L_y \). Given that one is trying to recall items from \( L_x \), there will be an item-to-context association for each \( L_x \) image and the context that is used as the retrieval cue (\( a_{Lx} \)) and an item-to-context association for each \( L_y \) image and context used to probe memory (\( a_{Ly} \)). On these assumptions, the probability of sampling image \( I \) from \( L_x \) is:

\[
P(I_{Lx},Q) = \frac{a_{Lx}}{\sum_{j=1}^{m} a_{Lx} + \sum_{j=1}^{m} a_{Ly}}, \quad (1)
\]

and the probability of mistakenly sampling image \( I \) from \( L_y \) is:

\[
P(I_{Ly},Q) = \frac{a_{Ly}}{\sum_{j=1}^{m} a_{Lx} + \sum_{j=1}^{m} a_{Ly}}. \quad (2)
\]

Once an image has been sampled from memory a recovery of the contents of that image is attempted, which is successful with following probability:

\[
R(I_n,Q) = 1 - \exp(-S(I_n,Q)) = 1 - \exp(-a_{Lx}), \quad (3)
\]

where \( n = x \) or \( y \). Thus, the product of Equations 1 and 3 give the probability of successfully recalling a given item from the target list. In contrast, the product of Equations 2 and 3 give the probability of recalling a given item from a non-target list (i.e., an intrusion error).
implemented by assuming that more recent lists should be better recalled than less recent lists. This is because the strength of the context-to-image association differs between images on different lists (Mensink & Raaijmakers, 1989). Thus, the strength of the association between the test context used and the items in memory is greater for more recent lists. Consider Figure 1. Assume that one is trying to recall $L_x$. Instructions to forget $L_y$ that is increasing $a_{L_y}/a_{L_x}$, causes an increase in the probability of recalling an $L_y$ item. This is the benefit of directed forgetting. Now assume that one is trying to recall $L_y$. Instructions to forget $L_x$, causes a decrease in the probability of recalling an $L_x$ item. This is the cost of directed forgetting.

**Predictions**

The SAM model makes two novel predictions. First, the more recent list should be remembered better than the less recent list because the test context used to probe memory is assumed to be more similar to the more recent list (cf. Mensink & Raaijmakers, 1989). This is a critical prediction of the model. A review of directed forgetting literature shows, however, the opposite is almost always the case.

We note that many of the experiments in the literature confound a number of variables. Because the list method usually makes use of only two lists, $L_1$ and $L_2$, the effect of $L_1$ versus $L_2$ is confounded with presence of an interfering prior list. In addition, subjects usually do not perform a distractor task after $L_2$. Lack of a subsequent distractor task benefits $L_1$ because the last items on $L_1$ may be rehearsed during $L_2$ (Peterson & Peterson, 1959; Rundus, 1971). Lastly, subjects are typically asked recall both $L_1$ and $L_2$ at test simultaneously, which makes it somewhat plausible that output interference explains directed forgetting.

To control for these variables, we used a three-list design in the present experiment (cf. Sahakyan, 2004). Thus, both $L_2$ and $L_3$ were preceded by a prior list. (Memory for $L_1$ was never tested.) With respect to Figure 1, $L_2 = L_x$ and $L_3 = L_y$ in the present experiment. The distractor task is traditionally used as a means for controlling rehearsals (Peterson & Peterson, 1959). Thus, a distractor task was performed after each list, and the number of rehearsals for different lists should be controlled. Lastly, subjects recalled one list at a time in order to control for output interference; those subjects asked to recall $L_2$ can do so when not also attempting to recall $L_3$ items (e.g., Sahakyan & Kelley, 2002). With these controls in place, the simplest versions of the rehearsal and output interference accounts of directed forgetting predict no effect of directed forgetting in the present experiment, whereas the CD account does.

The second prediction concerns intrusion errors. Intrusion errors are notoriously difficult to investigate because so few are made. Nevertheless, the assumption that the test context used to probe memory is more similar to the $L_3$ context than to the $L_1$ context leads to the prediction that the number of $L_3$ intrusion errors when trying to recall $L_2$ should be greater than the number of $L_2$ intrusion errors.
when trying to recall $L_3$. This difference should increase as the result of instruction to forget (i.e., increasing $a_{L_3}/a_{L_3}$).

Figure 1 shows that instructions to forget $L_n$ produce fewer $L_n$ intrusions when attempting to recall $L_n$. That is increasing $a_{L_3}/a_{L_3}$ decreases $L_n$ intrusion errors, which are shown on the dashed line below the horizontal line in Figure 1. In contrast, instructions to forget $L_n$ produce more $L_n$ intrusions when one is attempting to recall $L_n$. Increasing $a_{L_3}/a_{L_3}$ increases the number of $L_n$ intrusion errors.

We also anticipate that some $L_3$ items will be recalled and output erroneously. The tendency to intrude an $L$ item is predicted to occur more often when trying to recall $L_2$ than $L_3$ because the context used to probe memory will be more similar to $L_1$. In addition, the model predicts fewer $L_3$ intrusions after forget instructions because the cue used to probe memory at test should be less similar to the $L_3$ context, regardless of whether $L_2$ or $L_3$ is to be recalled. In the next section, we report the results of an experiment designed to test these predictions.

Method

Subjects. 165 undergraduate students from the University of South Florida were randomly assigned to each condition.

Design, Materials, and Procedure. For each subject, 48 nouns were randomly selected from Francis and Kucera (1982) corpus with frequencies between 20 and 50 occurrences per million and randomly divided into three study lists. Subjects were test individually in a sound-attenuated subject booth containing a personal computer. Each word was presented on the computer monitor and studied for 8 s. After each study list subjects performed a 30 s. distractor task, which involved mentally adding single digits.

A 2 (remember versus forget $L_3$) x 2 (recall $L_2$ versus $L_3$ at test) between subject design was used. Following the $L_2$ distractor task 83 subjects (forget group) were informed that memory would be tested for $L_3$. The assumption was that informing subjects that $L_3$ would be tested would cause subject to attempt to forget $L_2$. Of these, 41 subjects ($L_2$ forget) were then informed after the $L_3$ distractor task that their memory for $L_2$ would actually be tested and 42 ($L_3$ forget) were informed that their memory $L_3$ would be tested.

Eighty-two subjects (remember group) were not informed after $L_2$ that there memory for $L_3$ would be tested. Rather, 42 of these subjects were told after the $L_3$ distractor task that only their memory for $L_2$ ($L_2$ remember) would be tested and 40 ($L_3$ remember) were told that only their memory for $L_3$ would be tested.¹ When memory was tested the subjects had one minute to recall as many words from the to-be-recalled list as they could. They were free to enter their responses in order they chose. Subjects used a computer keyboard to enter their responses.

Results

An alpha = .05 is adopted as the standard of significance. Figure 2 shows a reliable interaction between list number and the forgetting instructions [$F(1, 161) = 19.7$]. In addition, there was a reliable recency effect in the remember condition [$F(1,80) = 5.6$], which is consistent with the assumption that the test context is more similar to $L_3$ when trying recall $L_3$ than it is to $L_2$ when trying to recall $L_2$. Table 1 lists the mean intrusion rates. Here there were

![Figure 2. SAM model fits to the correct recall data. The parameters used to fit the model were: $a_{L3L2R} = a_{L3L2F} = .42$; $a_{L2L3R} = .35$; $a_{L2L3F} = .18$.](image)

Table 1. SAM model fit to the intrusion data.

<table>
<thead>
<tr>
<th>Intrusion List</th>
<th>Recall List 2 Data</th>
<th>Recall List 2 Fit</th>
<th>Recall List 3 Data</th>
<th>Recall List 3 Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remember</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List 1</td>
<td>.051</td>
<td>.054</td>
<td>.016</td>
<td>.017</td>
</tr>
<tr>
<td>List 2</td>
<td>NA</td>
<td>.019</td>
<td>.023</td>
<td></td>
</tr>
<tr>
<td>List 3</td>
<td>.030</td>
<td>.036</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Forget</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List 1</td>
<td>.038</td>
<td>.039</td>
<td>.010</td>
<td>&lt;.010</td>
</tr>
<tr>
<td>List 2</td>
<td>NA</td>
<td>.015</td>
<td>&lt;.010</td>
<td></td>
</tr>
<tr>
<td>List 3</td>
<td>.023</td>
<td>.023</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

¹ A final round of testing occurred, whereby subjects who were asked initially to recall $L_2$ were asked recall $L_3$ and vice versa. We defer discussion of these data for later.
reliably fewer $L_j$ intrusions for $L_3$ than for $L_2$ \[ F(1, 161) = 17.8 \]. None of the other main effects or the interaction reliably affected intrusion rates.

**Model Fitting**

To fit the model to the data, we obtained a set of $a$ parameters that provided a reasonable fit to the correct recall data (Figure 2). Call $a_{i,j,1,2}$ the strength of the context-cue to $L_j$ image strength when attempting to recall $L_j$ items in the remember condition ($R$) and $a_{i,2,2,2}$ the strength of the context-cue to $L_2$ image strength when attempting to recall $L_2$ items. Likewise, replace the $R$ with an $F$ in preceding notation for the forget condition. We assumed that the forget instruction did not affect the test-context to $L_3$ image strength. Hence, $a_{i,3,1,3} = a_{i,3,1,3}$.

All the parameter values used in the fit are listed in Figure 2. The model fits the data well and the $a$ values are consistent with the assumptions that the test context is more similar to $L_1$ than to $L_2$ and that the forget instruction produced an increase in $a_{i,3,1,3}$. We also generated fits for the intrusion errors (Table 1). In this case, we were not able to make any a priori assumptions other than the context-to-item strengths should lower for those lists that are not to be recalled. Thus, this fit is not necessarily a test of the model, but it does allow us to interpret the data within the framework of the SAM model.

The pattern of intrusion data suggests that the context used to probe memory is more similar to $L_1$ when $L_2$ is tested than when $L_1$ is tested ($a_{i,1,1,2} > a_{i,1,1,3}$). In other words, the $L_1$ and $L_2$ context are relatively similar, which one would expect based on the SDSS assumptions. In addition, the instructions to forget produced lower context to $L_1$ strength estimates, which is consistent with the hypothesis that instructions to forget produce a larger change in context than when the instruction to forget is not given. A similar conclusion is drawn based on the $L_2$ intrusions when $L_1$ was to be recalled (i.e., $a_{i,2,1,3} > a_{i,2,1,2}$).

The present set of CD assumptions comes up short in one respect, however. We assumed that the context used to probe memory for $L_2$ items is more similar to $L_1$ after instructions to forget. If so, the predictions is that the $L_1$ intrusions should be greater following instructions to forget, but they were not reliably different from the remember condition. In fact, the model fit suggests that context used to probe for $L_2$ is less similar to $L_1$ after instructions to forget, which is inconsistent with CD.

We have two reactions. First, intrusions rates are difficult to investigate because intrusions are so infrequent, and thus some would opine that that they are not worth trying to account for at all. However, some of the patterns of intrusions rates are reliable and sometimes the intrusion rates were nearly as large as the rate of correct responses. Theoretically speaking, the finding that the forget instruction does not increase the rate of $L_3$ intrusions when trying to recall $L_2$ suggests that there might be an evaluation of the source from which an item was sampled and recovered (cf. Raaijmakers & Shiffrin, 1980), and increasing the strength of association between the test context to $L_3$ images might allow $L_3$ items to be more effectively discriminated from the $L_2$ images during this source evaluation process. At this point, however, we defer an extension of the model until more data are collected, and we conclude that the model captures all of the major trends in the data, and the model should not be rejected.

**General Discussion**

In this paper, we have presented the first formal model of the suppression, repression, or inhibition involved in directed or intentional forgetting. The critical assumption is that directed forgetting is a context effect. When one attempts to forget a recent event, one attempts to “think of something else” and the result is a greater change in cognitive context than what one would expect without the intention to forget (Sabakyan & Kelley, 2002).

The SAM model captures the major trends in the data without appealing to the rehearsal or output interference explanations. That, in and of itself, does not necessarily mean that the rehearsal and output interference accounts are incorrect. However, the results of the present experiment do challenge them on empirical grounds. The present experiment controlled for rehearsals by imposing a distractor task after each list, which should have eliminated the effect of forget instructions according to the rehearsal account. In addition, output interference was controlled, which should have eliminated the effect of the forget instructions according to the output interference account.

It is perhaps interesting or even important to note that in fact our subjects were not directed in any explicit sense to forget anything. Rather, subjects were simply informed after studying two lists that their memory would be tested for the next list. The result was a large cost associated with remembering the list that was studied prior to the instructions and a large benefit associated with the to-be-remembered list. Thus, it appears that intentional forgetting is a part of the metacognitive repertoire of the average subject, suggesting that intentional forgetting is a more general strategy for reducing interference for relatively important events, albeit at the expense of a lower rate of remembering the immediately preceding events.

**Context Differentiation during Rehearsal**

While the current SAM model provides an explanation of directed forgetting based on the assumption that the intention of forgetting produces a change in cognitive context, we have not shown how it can account for directed forgetting using the item-method. As it turns out, it is straightforward to do so, and it too is assumed to be a context effect.

Recall that the item method instructs subjects after the presentation of each item on a list whether memory for that item will be tested. In this case, we also assume that the subject attempts to forget those items for which memory will not be tested by “thinking of something else” or changing the mental context. There is, however, a critical
difference between the study conditions associated with the list and item methods.

In the list method, it makes sense to not think about any of the items on prior list because these items are to be forgotten. Thus, the subject might think about what they are going to do after the experiment is over, for instance. In the item method, on the other hand, it makes sense to think about the items that are to be remembered. Therefore, subjects might employ a strategy of thinking about or rehearsing a recent item that is to be remembered following an instruction to forget the current item (e.g., Bjork & Geiselman, 1978). This amounts to covert spaced repetitions of the to-be-remembered items, which will increase the amount or strength of context storage (Malmberg & Shiffrin, 2005).

SAM (Raaijmakers & Shiffrin, 1980) assumes that the context-to-item strength ($a$) is linearly related to the number of times ($t$) an item is rehearsed: at. Malmberg and Shiffrin (2005) qualified this assumption to apply only to spaced repetitions. Thus, to-be-remembered items will have a greater average context-to-item strength than to-be-forgotten words, whereas the model of the list method assumes that $t$ is the same for all items. According to Equation 1, therefore, instructions to forget will increase the probabilities that a to-be-remembered item will be sampled and a to-be-forgotten item will not be sampled.

Intuitively speaking, these assumptions are in accord with the CD assumption that intention to forget produces a differentiation in memory of items that are to be remembered from those that are to be forgotten, and thus provides a parsimonious explanation of the list and item methods of direct forgetting. A theoretical limitation of the SAM model in its present form is that it assumes that context changes between lists more robustly as the result of intentions to forget, but it does not describe how these changes take place.

As a member of the SAM family of models, it seems most parsimonious to adopt a set of assumptions about contextual fluctuation that were described by Mensink and Raaijmakers (1989). Accordingly, context randomly fluctuates over time, and the similarity between the test context and the context stored during study decreases exponentially with increases in time or the number of intervening lists. The advantage of this model is that it reduces the number of free parameters because the $a$ parameters associated with $L_1$, $L_2$, and $L_3$ are systematically related. No additional assumptions are required to implement the CD assumption in the Mensink and Raaijmakers (1989) for the item method; we can assume that to-be-remembered items have stronger context-to-item associations because they are preferentially rehearsed. For the list method, one could assume that instructions to forget cause a perturbation in the contextual elements active after the to-be-forgotten list that is larger than would be the case if instructions to forget were not given. Our aim is to report such a model in near the future.

References


