Psychological Science OnlineFirst, published on January 17, 2012 as doi:10.1177/0956797611430692

Research Report

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
XX(X) 1–5

DOI: 10.1177/0956797611430692
http://pss.sagepub.com
SAGE

Overcoming the Negative Consequences of Interference From Recognition Memory Testing

Kenneth J. Malmberg[1], Amy H. Criss[2], Tarun H. Gangwani[3], and Richard M. Shiffrin[3]
[1]University of South Florida, [2]Syracuse University, and [3]Indiana University, Bloomington

Abstract

Theories of why humans forget have been challenged by the newly discovered list-length/output-interference paradox, in which—under certain testing conditions—learning is not harmed by the amount of verbal material studied, whereas retrieval of that material becomes more difficult with increases in the number of items tested. The latter finding is known as output interference, and the results of the experiment reported here indicate that a release from output interference is obtained when the nature of the items is changed during testing. Specifically, when participants are asked to recognize items from two categories, output interference is minimized when items from each category are tested separately in large blocks. This finding supports models of forgetting that assume interference arises from information about the to-be-learned material that is stored in memory; in contrast, this finding is difficult to explain using models that assume forgetting is the result only of changing context.

Keywords

forgetting, memory

Received 5/27/11; Revision accepted 9/3/11

Recognition refers to people's ability to determine what they have previously experienced. For instance, an individual might be asked whether a certain person was present at a crime scene (i.e., whether a specific item was represented in a specific context). The ability to accurately recognize past events characterizes healthy memory, and deficits in recognition memory are hallmarks of memory disorders, such as Alzheimer's disease (Balota, Burgess, Cortese, & Adams, 2002). Not surprisingly, recognition has been extensively researched. In such research, participants are usually presented with a list of words and asked to judge which words were presented on a study list and which were not (see Malmberg, 2008, for a review). A key finding of such studies is that increasing the number of items tested decreases the accuracy of both yes/no responses and forced-choice recognition; this decrease is referred to as *output interference* (Criss, Malmberg, & Shiffrin, 2011).

Output interference has been extensively documented in free-recall tasks (e.g., Raaijmakers & Shiffrin, 1981; Roediger, 1974; Roediger & Schmidt, 1980; Tulving & Arbuckle, 1966), but it has received little attention in recognition-memory studies. Output interference is similar to *proactive interference*, in which the encoding of prior events negatively affects the ability to remember subsequent events. Nevertheless, output interference in recognition testing may at first glance be surprising given the reports that testing memory enhances learning (e.g., Karpicke & Roediger, 2008) and given that practice usually enhances human performance (e.g., Anderson, 1982; Shiffrin & Schneider, 1977). In the experiment reported here, we focused on output interference due to the testing of foils and previously tested targets. We did this by exploring conditions in which this memory limitation can be overcome.

Output Interference and the Release From Proactive Interference

One way to overcome output interference is suggested in the well-established literature on proactive interference. In one influential series of experiments, Brown (1958) and Peterson and Peterson (1959) reported that recall for a trigram declined

Corresponding Author:
Kenneth J. Malmberg, Department of Psychology, University of South Florida, PCD 4118G, Tampa, FL 33620
E-mail: malmberg@usf.edu

as both the duration of distracting activity and the number of test trials increased (Keppel & Underwood, 1962). Watkins and Watkins (1975) attributed the latter result to an increase in proactive interference. Wickens, Born, and Allen (1963; see also Wickens, 1970) demonstrated that memory for stimuli is enhanced when the type of stimuli being tested is changed. For example, Wickens and his colleagues had participants study different categories of trigrams; on subsequent recall tests, performance decreased over the initial trials, which focused on one category of trigrams, but performance rose to a level near that for the first trial after the second category was introduced. This phenomenon is termed *release from proactive interference*, and these results suggest that a switch in the class of stimuli during recognition testing may also cause a release from output interference.

The Nature of Forgetting

Models of memory and forgetting assume that interference arises at test from traces of the items that have been studied (*item noise*) as well as from contextual or source confusions resulting from representations of the test items prior to study (*context noise*). *Item-noise models* provide straightforward accounts for several findings indicating that the nature of the words studied and the composition of the study lists have an impact on recognition (e.g., Criss, 2006; Criss & Shiffrin, 2004; Malmberg & Murnane, 2002; Malmberg, Steyvers, Stephens, & Shiffrin, 2002); item-noise models predict output interference on the assumption that memory traces are stored for both targets and foils during testing (Criss et al., 2011).

For instance, in a pure item-noise version of the retrieving-effectively-from-memory theory (Shiffrin & Steyvers, 1997), each study trial is represented by a set of features representing the studied word and the context in which it occurred. At test, context features are used to isolate the relevant subset of memory for the list on which the word appeared, and item features of the test item are compared with the traces stored during study and test. Each such match contributes to the familiarity of the test item, with familiarity increasing as a function of how similar the test item is to the stored traces. For example, both targets and foils are more likely to be judged old as the number of studied items from the test item's category increases (Criss & Shiffrin, 2004). Storage during test results in either a new memory trace (i.e., if the test item is judged new) or the updating of a stored memory trace (i.e., if the test item is judged old). Increasing the number of foils increases the number of traces that do not match a subsequent test item. This added noise reduces recognition performance. Furthermore, the amount of noise and corresponding reduction in recognition performance vary depending on the similarity of successive test items.

One implication of the item-noise assumption is that output interference is related to the number of items tested, not necessarily to the number of discrete test trials. Murdock and Anderson (1975) found that recognition accuracy decreased as the number of alternatives in forced-choice testing increased. Specifically, as Figure 1 shows, accuracy on two-alternative forced-choice (2AFC) tests decreased approximately 3% across test trials, but accuracy on 6AFC tests decreased 17%; these results suggest that items, not trials or time, are the source of output interference.

According to a different class of models (e.g., the bind-cue-decide model of episodic memory, Dennis & Humphreys, 2001), recognition performance depends on the number and variety of different contexts in which verbal material has been previously encountered. Such models predict that the more contexts in which a word has occurred, the more difficult it is to determine whether it occurred in any specific context, such as in a study list. *Context-noise models* assume that in recognition studies using words, only context noise operates and that forgetting is due solely to confusions arising from the prior contexts in which a word has appeared. According to context-noise models, other items do not contribute to recognition.

Context-noise models can explain output interference only with additional assumptions, such as changing context during testing. For instance, context-noise models may assume that output interference is caused by a decline in the ability to mentally reinstate the study context across the course of testing. Evidence casting doubt on this interpretation is provided by our previous finding that the length of the interval between study and test does not affect the amount of output interference (Criss et al., 2011). In that study, the decline in memory

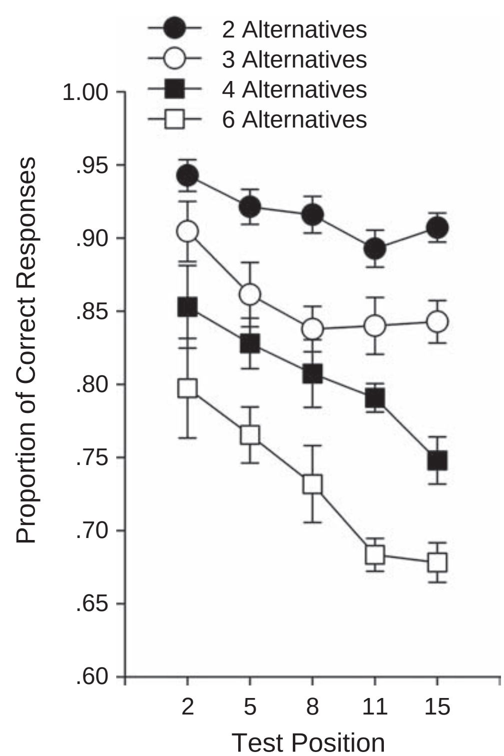

Fig. 1. Results from Murdock and Anderson (1975): mean proportion of correctly recalled study items as a function of test position and the number of forced-choice response items. Error bars show standard errors.

due to about 3 min of testing was far greater than the decline following a 20-min increase in the interval between study and test.

Experiment

The goal of our experiment was to demonstrate a release in output interference and show that changes in the nature of the items being tested determine recognition accuracy. We used a novel design that manipulated the composition of the test list but held study conditions constant (e.g., Neely & Tse, 2008). Items from two categories were randomly intermixed during study. Participants were then tested in 150 trials with a 2AFC design. In one condition, the order in which the words were tested was determined randomly. In two other conditions, testing was blocked by category. The items within a category were more similar than items belonging to different categories, at least on average. Therefore, according to item-noise models, repeated testing from the same category should increase noise and decrease accuracy. Further, switching the item information used to probe memory should decrease the item noise generated by the prior encoding of tested targets and foils, and thus a release from output interference should be observed in the form of an increase in accuracy.

Method

Participants. Eighty-eight undergraduate students at Indiana University earned course credit for participating in the experiment. Participants were randomly assigned to three conditions: 27 to the random condition, 36 to the blocked condition, and 25 to the short-block condition.

Materials. The stimuli consisted of 1,200 English words grouped into four sets of 300 words each. The words in each set were split into two 150-word categories and grouped as follows: animals/geological terms, adjectives/foods, proper names/company names, and countries/professions. Some brand names contained hyphens (e.g., Wal-Mart) or an apostrophe (e.g., Kellogg's), but no words in any category contained spaces. For each participant, half of the words from each category were randomly assigned as targets, and the other half were assigned as foils.

Procedure and design. Participants completed one study-test cycle for each of the four stimulus sets. Sets and stimuli were presented in a random order. Each cycle began with 150 study trials: 75 target words from each of the two categories randomly intermixed. On each study trial, a word was presented in the center of a computer screen for 1,000 ms, followed by a 100-ms interstimulus interval. The study trials were followed by 150 test trials with a 2AFC design. Each test trial contained two items from the same category, the target and a foil (one of the 75 nonstudied words from that category), which were randomly paired. On each self-paced test trial, the target and foil were presented side by side in the center of the computer screen, with the target location (i.e., left or right) randomly chosen. Participants were instructed to indicate which of the two words had been shown in the study trials.

The study conditions were identical for all participants, but test conditions varied. In the random condition, the order of test trials was randomized separately for each subject. In the blocked condition, 75 trials presenting all of the words from one category preceded 75 trials presenting all of the words from the other category. The transition between categories was seamless, and subjects were not forewarned. In the short-block condition, the category type alternated every five test trials. The order of category testing was randomized separately for each subject.

Results

Results for all conditions showed substantial output interference. The data from each 150-trial test list was divided into 30 five-trial blocks. Performance on these blocks was subjected to a 3 (test condition; between subjects) × 30 (block; within subjects) mixed-factor repeated measures analysis of variance. As Figure 2 shows, we found a significant main effect of block, $F(29, 2465) = 23.65$, $MSE = 0.01$, $\eta_p^2 = .22$, $p < .0005$, but not of test condition, $F < 1$. In all test conditions, recognition accuracy decreased as the number of test blocks increased (all Fs > 100 and ps < .0005).

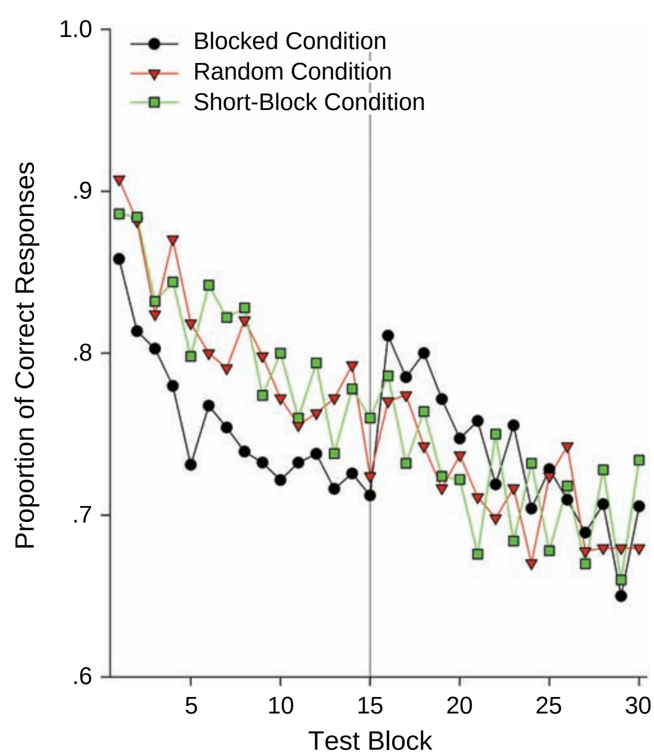

Fig. 2. Results from the current experiment: mean proportion of correctly recalled study items as a function of test block and condition. The vertical line indicates the point at which test categories were switched in the blocked condition.

Only participants in the blocked condition demonstrated release from output interference. There was a reliable condition-by-block interaction, $F(58, 2465) = 2.89$, $MSE = 0.03$, $\eta_p^2 = .06$, $p < .0005$. As Figure 2 shows, switching the category after 15 blocks of testing produced a release from proactive interference of approximately 10%. However, the release was not sufficiently large to equate accuracy on Block 1 and Block 16, $t(36) = 6.15$, $SEM = 0.01$, $p < .0005$, which indicates that the release from output interference was incomplete. In contrast, the same analyses for the short-block condition revealed no reliable effect, which indicates that brief category switching in and of itself is not sufficient to induce a significant release from proactive interference.

To further analyze the release from proactive interference, we compared accuracy on the 25 trials before and after the midpoint of the list. Again, there was a significant block-by-test-order interaction, $F(2, 86) = 12.88$, $MSE = 0.002$, $\eta_p^2 = .23$, $p < .0005$, which indicated that accuracy improved only in the blocked condition, $t(37) = 4.72$, $SEM = 0.01$, $p < .0005$.

Discussion

This experiment demonstrated that substantial output interference can occur on a 2AFC recall task, but it showed that changing the nature of the test items can significantly reduce such interference. One lesson to take from these findings is that extended testing of like items will yield considerable underestimation of the quality of recognition memory. More accurate assessments of what has been stored in a given context, and what can be retrieved from that context, can be obtained by conducting only a small number of tests on a given class of stimuli. Overtesting memory for stimuli of like kinds results in a buildup of proactive interference that increasingly impedes the ability to remember items as testing proceeds.

Because a release from output interference was not observed in the short-block condition, it is clear that there is a limit on the degree to which switching testing between item classes can enhance performance. When test blocks consist of only five items, release from output interference should be minimal because many prior test traces will be similar to any item tested. It is less clear how long a block must be in order to observe a release from output interference. There are likely costs associated with switching the contents of retrieval cues, say from emphasizing one set of item features representing category membership rather than another set. The costs of switching cues may balance the costs of output interference, and only long blocks may tip the balance toward domination by output interference. The results of the short-block condition may also be informative about the duration over which release from proactive interference occurs.

Early studies tested memory immediately or after a very short delay. For example, Kincaid and Wickens (1970; also see Neely, Schmidt, & Roediger, 1983) showed release from proactive interference by waiting 120 s between the three interference trials and the fourth trial, even when the fourth trial contained three items from the same category. In the experiment reported here, we observed no release from output interference over the course of five test trials from a different category. Further research is needed to define optimal conditions for observing the release from output interference in recognition-memory tasks.

The observation that a release from output interference can be achieved via a switch in the testing materials strongly implicates the role of item noise as a factor that limits the level of recognition accuracy. Although this result is predicted by item-noise models (Criss et al., 2011), it is difficult to explain using context-noise models. The problem posed for context-noise models is that output interference negatively affected performance over at least the first 15 blocks of testing in all conditions; in addition, the release occurred when a different category of test items was introduced. According to context-noise models, the similarity among items stored in memory should not affect recognition performance because these models predict no interference from other items. Perhaps extra assumptions could be added to context-noise models to explain findings such as the ones reported here, but item-noise models provide a relatively simple and straightforward explanation of our results.

Output interference is also one part of mnemonic paradox. Although many reports have concluded that increasing the number of items studied harms recognition (e.g., Murnane & Shiffrin, 1991; Strong, 1912), Dennis and his colleagues found that recognition memory for words was independent of the number of words studied (Dennis & Humphreys, 2001; Dennis, Lee, & Kinnell, 2008); this finding is called the *null list-length effect*. Although these recent reports, of course, are controversial, and the size of list-length effects remains a topic of debate, the combination of relatively small list-length effects and relatively large effects of output interference presents a challenge to all extant models of memory. However, any viable explanation will have to take into account the release from output interference that we observed here when stimulus categories were switched.

Acknowledgments

K. J. M. and A. H. C. contributed equally to this report.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

Funding for this study was provided by a grant from the National Science Foundation (0951612) to A. H. C. and by a grant from the Air Force Office of Scientific Research (AFOSR-BAA-2010-1) to R. M. S.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). Memory for the infrequent in young, old, and early stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, *46*, 199–226.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*, 12–21.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*, 461–478.

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*, 316–326.

Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis & Humphreys (2001). *Psychological Review*, *111*, 800–807.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–478.

Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list-length effect. *Journal of Memory and Language*, *59*, 361–376.

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*, 153–161.

Kincaid, J. P., & Wickens, D. D. (1970). Temporal gradient of release from proactive inhibition. *Journal of Experimental Psychology*, *86*, 313–316.

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335–384.

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 616–630.

Malmberg, K. J., Steyvers, M., Stephens, J., & Shiffrin, R. (2002). Feature frequency effects in recognition memory. *Memory & Cognition*, *30*, 607–613.

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage and retrieval of item information. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 145–194). Hillsdale, NJ: Erlbaum.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 855–874.

Neely, J. H., Schmidt, S. R., & Roediger, H. L. (1983). Inhibition from related primes in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 196–211.

Neely, J. H., & Tse, C.-S. (2008). Semantic relatedness effects on true and false memories in episodic recognition: A methodological and empirical review. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 313–352). New York, NY: Psychology Press.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198.

Raaijmakers, J., & Shiffrin, R. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*, 261–269.

Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired-associate lists. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 91–105.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review*, *19*, 447–462.

Tulving, E., & Arbuckle, T. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology*, *72*, 145–150.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *104*, 442–452.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, *77*, 1–15.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *2*, 440–445.