QJEP

Release from output interference in recognition memory: A test of the attention hypothesis

⑤SAGE

Amy H Criss[1], Cristina Salomão[1], Kenneth J Malmberg[2], William Aue[1,3], Aslı Kılıç[1,4] and MarkAvery Claridge[1]

Abstract

Retrieval results in both costs and benefits to episodic memory. Output interference (OI) refers to the finding that episodic memory accuracy decreases with increasing test trials. Release from OI is the restoration of original accuracy at some point during the test. For example, a release from OI in recognition memory testing occurs when the semantic similarity between stimuli decreases midway through testing, suggesting that item representations stored on early trials cause interference on tests occurring on later trials to the extent that the earlier items share features with the latter items. In two recognition memory experiments, we demonstrate release from OI for words and faces. We also test whether release from OI is the result of interference or is due to a boost in attention caused by reorienting to a novel stimulus type. A test for the foils presented during the initial test list supports the interference account of OI. Implications for models of memory are discussed.

Retrieval both helps and harms episodic memory. For example, long-term memory can benefit more from testing memory than from additional study (e.g., Roediger & Karpicke, 2006a, 2006b) However, the benefits of testing are far from ubiquitous, and memory is even harmed by testing in some cases (Malmberg, Lehman, Annis, Criss, & Shiffrin, 2014). For instance, accuracy decreases across a series of episodic memory test trials, known as output interference (OI, for example, Criss, Malmberg, & Shiffrin, 2011; Kilic, Criss, Malmberg, & Shiffrin, 2017; Koop, Criss, & Malmberg, 2015; Murdock & Anderson, 1975), but some situations buffer against OI (Malmberg, Criss, Gangwani, & Shiffrin, 2012; Watkins & Watkins, 1975). Identifying the conditions under which OI occurs and under which release from OI is observed in long-term memory is surprisingly under investigated and is, thus, the focus of this article.

In a typical recognition memory paradigm, participants study a series of items, and after some delay, they are asked to discriminate between studied targets and unstudied foils. There are many widely used recognition procedures, including single-item recognition and forced-choice recognition. Whereas single-item recognition is ideal for some

research questions, especially those where understanding the nature of false memory is critical (e.g., Criss & Shiffrin, 2004a; Gallo, 2010; Roediger & McDermott, 1995), forced choice is ideal for research questions regarding accuracy in the absence of response bias (see Green & Swets, 1966; Grider & Malmberg, 2008; Lockhart & Murdock, 1970 cf., Starns, Chen, & Staub, 2017). In single-item recognition, performance is typically characterized by examining the probability of correctly identifying a target as studied (hit rate, HR) and the probability of incorrectly endorsing a foil as studied (false-alarm rate [FAR]). The pattern of performance across the test trial differs for HR and FAR. HRs decreases substantially as a function of test trial, whereas the pattern of FARs across the test trial is less clear but

[1]Syracuse University, Syracuse, NY, USA
[2]University of South Florida, Tampa, FL, USA
[3]Purdue University, West Lafayette, IN, USA
[4]Middle Eastern Technical University, Ankara, Turkey

Corresponding author:
Amy H Criss, Department of Psychology, Syracuse University, Syracuse, NY 13224, USA.
Email: amy.criss@gmail.com

tends to be relatively flat (see Criss et al., 2011; Koop et al., 2015). The combined result is a decrease in summary measures of recognition accuracy (e.g., d', A') with increases in the number of trials tested. Correspondingly, in forced choice, OI presents as a decrease in accuracy (e.g., percent correct) with increasing test trial (Malmberg, Criss, Gangwani, & Shiffrin, 2012; Murdock & Anderson, 1975).

Although OI is often quite substantial in recognition testing, there are cases where OI is absent. For one, recognition tests of semantic memory, in the form of classroom examinations or laboratory studies of general knowledge questions, do not suffer from OI (Aue, Criss, & Prince, 2015).[1] Interleaving semantic memory tasks within a series of recognition tests does not amplify OI beyond what is found from the episodic tasks alone (Annis et al., 2013). The focus of this article is another situation when OI is eliminated—when the class of stimuli being retrieved changes. In Malmberg et al. (2012), participants studied an intermixed list of words from two different semantic categories (e.g., produce and geological formations). The test was forced choice and was either blocked or mixed with respect to category order. Mixed testing resulted in OI, whereas a release from OI was observed when the category of test items changed in a blocked fashion. To our knowledge, this is the only report of release from OI in long-term recognition memory. Of course, the study of OI and release from OI has its roots in the release from proactive inhibition (PI) paradigm.

Wickens, Born, and Allen (1963) were among the first to demonstrate release from PI. In their release from PI design, participants studied trigrams from two different categories—either consonant strings or numbers. Each trigram was followed by a 10-s distractor task and then a prompt to recall the trigram. When sequential study-test trials presented the same type of stimuli, for example, a series of trials of consonant strings, recall decreased for each subsequent trial, referred to as PI. When the type of stimulus changed, recall performance rebounded to approximately the level of the first trial, this is referred to as a release from PI. Regardless of the trial position for the second category (4, 7, or 10 out of 10 total trials), participants showed release from PI for the second category.

Release from OI in long-term memory recognition memory differs from release from PI in multiple ways. The tasks encompass different time scales tapping potentially different memory systems (long vs short-term) and the ability to isolate the effect to testing. In OI, encoding and stimuli are both identical, only the order of the test differs. In PI, categorical information changes for both study and test which prevents a clear demarcation of the processes underlying the effect. Nevertheless, our theoretical understanding of OI is influenced by conceptualizations of PI. In a seminal paper, Wickens (1970) proposed that "the more psychologically similar the classes are, the more they will interfere with each other" (p. 3). He conceptualized PI as a means to identify features that characterize memory on the assumptions that the magnitude of PI was positively related to the similarity between items, and conversely, the magnitude of release from PI was negatively related to the similarity between stimuli. Wickens foreshadowed modern item-noise models of memory (e.g., Criss & Shiffrin, 2004a; Shiffrin & Steyvers, 1997) that predict interference due to shared features among items.

A recent model of OI is based on the same principles (Criss et al., 2010). In the model, items are similar by virtue of having similar features (either by chance or by design) and this overlap causes interference. Encoding items into memory during the test results in OI, and probing with dissimilar items during the test allow release from OI. Although the model's account of these findings is consistent with Wickens (1970) in that similar items cause interference, he raised the alternative hypothesis that the change in stimuli is obvious to the participant and causes her to "either learn it better or make a greater (and more successful) effort to retain this item than the previous one." In other words, release from PI may be the result of interference due to similarity or it may be the result of enhanced attention brought on by the obvious change in stimulus material. Within the PI literature, this enhancement is hypothesized to result in better encoding. In the OI design, the items are already encoded, and the change only happens during the test. Nevertheless, the same mechanism could cause heightened alertness and boost performance.

This mechanism has been proposed in contemporary models of memory that reject item interference as a cause of forgetting (e.g., Dennis & Humphreys, 2001; Kilic, 2012; Osth & Dennis, 2015, see Criss et al., 2010, for discussion). Context-noise models assume that the to-be-remembered context is compared to prior contexts associated with the test item. An item is remembered to the extent that the contexts match. No other items contribute to the decision, and thus, context-noise models predict no item interference. To explain OI, auxiliary assumptions must be adopted such as the waning attention hypothesis just described. Understanding the causal factors of OI and release from OI is critical not only for constraining theories of memory but also more generally for characterizing the features upon which episodic memory is built.

This article has three primary goals. First, a single study has demonstrated release from OI in long-term recognition memory (i.e., Malmberg et al., 2012). Given the theoretical and practical importance of this finding, it is important to replicate. Second, we evaluate whether OI and release from OI occur with words and faces. While a manipulation of semantic similarity resulted in a partial release from OI (Malmberg et al., 2012), single-item recognition does not place strong emphasis on semantic processing (Criss & Malmberg, 2008). Recognition memory is sensitive to changes in the form of the stimulus. For instance, in an

experiment in which either words or objects were studied, Criss and Malmberg (2008) observed substantial decreases in accuracy when subjects were tested with a different perceptual form of the items that were studied. Criss and Shiffrin (2004c, 2005) showed that memory for pairs (associative recognition) differed for word–word, word–face, and face–face pairs both in terms of overall accuracy and in the type of interference. Following Wickens' proposal that release from OI can provide clues to the sources of interference in memory, evaluating whether or not words and faces cause release is important to establish, especially if it results in a complete release from OI. Third, we ask to what extent release from OI is the result of recovered attention during the test? That is, is release from OI simply the result of a reorienting to the test after habituation to the stimuli or is it due to item interference as predicted by item-noise models? While this hypothesis is straightforward and widely invoked, we are not aware of any experiment evaluating the attention hypothesis of OI in long-term recognition, and thus, this investigation provides novel empirical and theoretical contributions.

Experiment 1

The purpose of this experiment is to evaluate OI and release from OI with words and faces as stimuli. Words and faces are encountered frequently in everyday life, and some analyses suggest that they are relatively well encoded during the study compared to infrequently encountered stimuli, like nonwords or Chinese characters (Xu & Malmberg, 2007). Superficially, words and faces would seem to consist of very different features (e.g., spatial and visual vs conceptual and verbal), and some neuroscience data (Haxby et al., 2001; Kanwisher, McDermott, & Chun, 1997) suggest a separate module or a differential distribution of neural activity responsible for processing faces. Therefore, a reasonable hypothesis derived from Wickens' (1970) theory is that the two classes of stimuli do not interfere with one another. On the other hand, words and faces are often associated with one another (e.g., names and descriptors). Viewing faces may bring to mind various words, causing interference if those semantic features are encoded in episodic memory traces and if those features are heavily weighted in a retrieval cue.

Participants studied a list of words, and unknown faces randomly intermixed and were tested with a mixed test or a blocked test. If words and faces share features, then OI should be observed in both cases due to the build-up of interference as test items are encoded into memory. In contrast, if words and faces are sufficiently dissimilar that they do not cause cross-category interference, then a release from OI should be observed in the blocked testing condition. Because the questions of interest are about accuracy and because people may have different response biases for words than faces, we use forced-choice testing.

Methods

Participants. The participants were 93 members of the Syracuse University research participation pool who earned either partial course credit or extra credit.

Materials. The word pool consisted of 800 high-frequency words ($M = 130.66$ from Kucera & Francis, 1967). The words varied in length between 4 and 11 letters. The face pool was that described in Criss and Shiffrin (2004c, 2005) and contains 210 grayscale photographs.

Procedure and design. Each participant received two study-test lists, one mixed test and one blocked test, in random order. The study list was composed in a similar fashion for both conditions: 50 words and 50 faces were randomly selected from their respective pools and randomly ordered for each participant. Each item was studied for 3 s followed by a blank screen for 500 ms. The study was followed by a 45-s distractor task of simple addition. The test list contained 100 two alternative forced-choice (2AFC) test trials where participants were instructed to select the studied target from two choices, including the target and a randomly selected foil from the same stimulus type. The test was self-paced and a 100-ms blank screen separated each trial.

All details of the study-test lists were identical, except the order of testing. In the mixed condition, word and face trials were randomly intermixed during the test. In the blocked condition, test trials were organized by stimulus type during the test, and the switch in category began at the midpoint of the test (i.e., Trial 51). Participants were randomly assigned to the order of faces then words or words and then faces.

Analysis plan

Analyses were conducted in JASP (Love et al., 2015). Frequentist statistics do not allow accepting the null hypothesis that two values do not differ. Therefore, when the null hypothesis is of theoretical importance, we will report Bayes Factors (BF) to quantify evidence in favor of the null (using default priors). The BF is the relative evidence for one of two competing models. In the case of a *t*-test, the relevant models are a null versus effect. In the case of an analysis of variance (ANOVA), we report comparisons of a model with main effects and no interaction to a model with main effects and an interaction. As reported here, a BF > 1 indicates evidence for null model. For example, a BF = 10 indicates that the data are 10 times more likely to be the outcome of a null effect than a model with an effect (see Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). BFs provide continuous evidence, and therefore we do not draw arbitrary labels, indicating that any value is "significant" or not (see Etz & Vandekerckhove, 2017; Morey, 2015). Note also that Bayesian analyses are sensitive to sample size and issues of

power (e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007).

Results and discussion

Accuracy for faces ($M=0.71$, standard deviation [SD]$=0.11$) and words ($M=0.69$, $SD=0.12$) did not differ and is not the critical comparison, and therefore we collapse over stimulus type, $t(92)=1.78, p=0.079, \text{BF}=1.935$. To evaluate OI, the test sequence was divided into 10 test bins each containing 10 trials, and we conducted a 10 (test bin) × 2 (test type) repeated-measures ANOVA. The results in Figure 1 are clear—performance decreases across the test in the both test lists, indicative of OI but there is a release from OI for the blocked condition. The ANOVA resulted in a main effect of test bin and an interaction between test bin and test type ($F(9, 828)=7.04, p<0.001$, $F(9, 774)=3.85, p<0.001$, respectively). Following Malmberg et al. (2012), we measure OI and release from OI by comparing performance on Test Bins 1 and 6. OI is indicated by a decrease in performance and full release from OI would be indicated by equivalent performance on Test Bins 1 and 6. There is an interaction between test bin and test type, $F(1, 92)=15.29, p<0.001$. The mixed list resulted in lower accuracy for Bin 1 than for Bin 6, $t(92)=4.72, p<0.001$, indicating OI. In contrast, blocking at test resulted in full release from OI, as accuracy was not worse in the sixth bin than the first bin of testing, $t(92)=-1.68, p=0.097, \text{BF}=2.276$. Another indicator of release from OI is better performance on Test Bin 6 than Bin 5, $t(92)=-3.06, p=0.003$) in the blocked condition. Accuracy did not differ from Bins 5 to 6 in the mixed list ($t(92)=1.64, p=0.104, \text{BF}=2.398$). Consistent with these individual tests, there is an interaction between test bin (1 and 6) and test type ($F(1, 92)=12.11, p<0.01$).

In summary, the switch in the class of stimuli caused a complete release from OI. The assertion is that there are a set of features that represent words that are not used to represent faces and vice versa in episodic memory traces and/or in the retrieval cues used to probe memory. Equally important, these results set the stage to allow a test of one hypothesized cause of the release from OI, specifically the potential role of waning attention when the nature of the stimuli changes during the test.

Experiment 2

In this experiment, we test whether the release from OI is the result of attention and enhanced encoding as reviewed by Wickens (1970) or the result of interference from similar features stored in memory. Participants studied a list of words and faces followed by two consecutive test lists. Test 1 included the same conditions as Experiment 1—mixed and blocked testing, and this serves as a replication. Test 1 was followed by Test 2, in
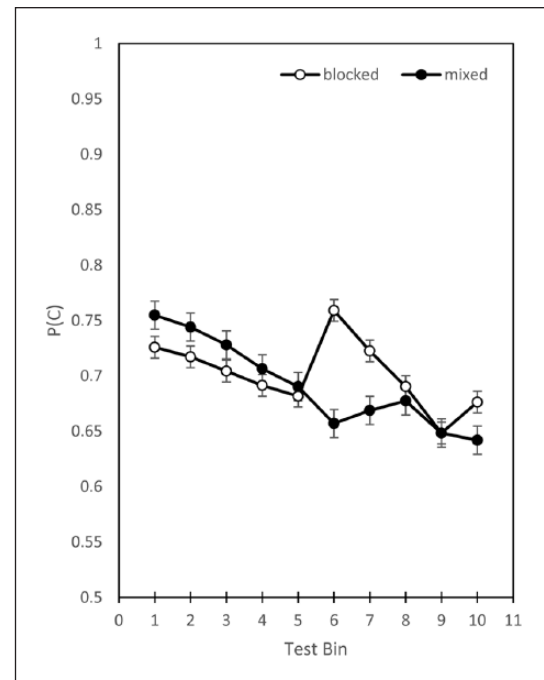

Figure 1. Results from Experiment 1 showing release from output interference when testing changes from words to faces or vice versa. Error bars represent ± 1 standard error.

which the participants were asked to discriminate foils from Test 1, which are now targets, from new foils. The purpose of Test 2 was to evaluate the attention hypothesis by evaluating whether memory for the foil items that occurred at and shortly after the switch point in Test 1 are better encoded. Suppose that the change in stimulus type draws attention, this would result in better encoding, effort or arousal, serving as the source of release from OI. Then, those first few items following the switch should be remembered better than other items from the test list. Figure 2 shows a simplified version of these predictions. If a change in stimulus category during Test 1 causes reorienting to the test and therefore better encoding, then the foils presented during those initial trials should be better remembered on a later test, Test 2 as shown in Figure 2a. If the category switch does not draw attention resulting in a release from OI and better encoding of the items following the Test 1 switch, then memory should be similar on average, subject of course to noise, as shown in Figure 2b.

Participants

The 143 participants were from the same pool described in Experiment 1.

Materials

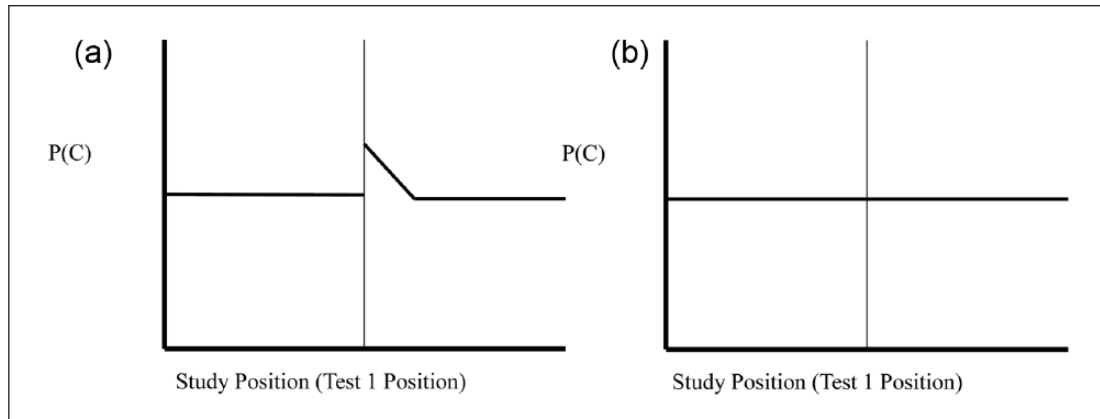The stimuli were those described in Experiment 1.

Figure 2. Stylized predictions for Experiment 2: (a) the attention hypothesis where the shift in stimulus category elicits reorienting and (b) no change in encoding due to the shift in category. The vertical bar represents the category switch.
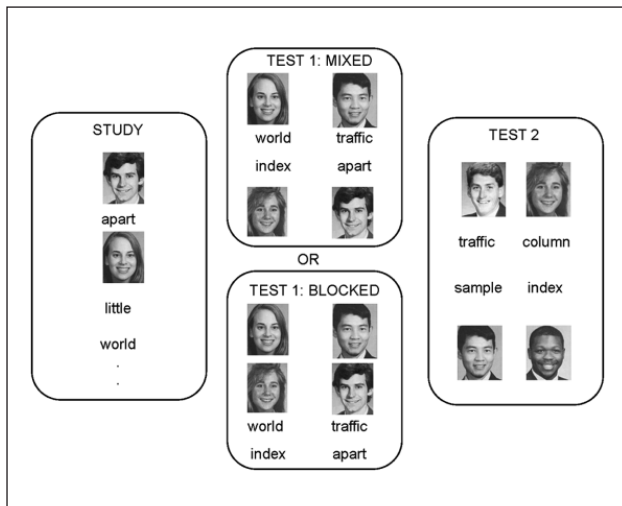

Figure 3. Design of Experiment 2. Each row indicates a trial. See further details in the main text.

Procedure and design

This experiment differed from Experiment 1 in two important details and is illustrated in Figure 3. First, participants received either a blocked or a mixed test list, but not both (conditions were run in different semesters). Second, the design was a study list followed by two consecutive test lists. The first study–test list was identical in methodology to either the blocked or the mixed list of Experiment 1. For the second test list, participants were asked to endorse foils from the first list and reject novel foils. In other words, the foils from Test 1 became targets on Test 2. Participants were not aware that there would be a second test. Test 2 was preceded by a distractor task identical in nature to the task separating study and Test 1, along with instructions so as to eliminate concerns about recency effects. Test 2 was mixed with respect to stimulus type for all participants, regardless of the condition to which they were assigned in Test 1.

Results and discussion

Test 1. As before, the test sequence was divided into 10 test bins each containing 10 trials. These data were submitted to a 10 (test bin) × 2 (test type) mixed ANOVA with test bin as a within-subject manipulation and test type as the between-subject manipulation. The results in Figure 4, left panel resemble those from Experiment 1. There was a main effect of test bin and an interaction between test bin and test type ($F(9, 1269) = 7.47$, $p < 0.001$; $F(9, 1269) = 2.75$, $p = 0.003$, respectively). The mixed list resulted in OI as measured by a decrease in accuracy from Test Bins 1 to 6, $t(66) = 3.94$, $p < 0.001$. The blocked test list resulted in a complete release from OI as indicated by no difference in accuracy between Test Bins 1 and 6, $t(75) = 0.05$, $p = 0.962$, BF = 7.91 and greater accuracy in Test Bin 5 compared to Bin 6, $t(75) = -2.18$, $p = 0.033$, compared to no difference in accuracy between Test Bins 5 and 6 in the mixed list, $t(66) = 0.94$, $p = 0.351$. There was an interaction between test bin and test type when comparing Test Bins 1 and 6 ($F(1, 141) = 7.22$, $p = 0.008$) and also when comparing Test Bins 5 and 6 ($F(1, 141) = 4.93$, $p = 0.028$). Fully replicating Experiment 1, OI was observed and a complete release from OI occurred when stimulus type changed.

Test 2. We now turn to the question of whether release from OI is facilitated by or the result of attention capture when the stimulus changes during the test. According to this hypothesis, the switch of the salient attributes of the test stimuli should attract additional attentional resources devoted to the encoding of the test stimuli. If so, during the second set of test trials, foils presented immediately following the switch in stimuli should be better recognized than those preceding the switch and better remembered than those tested in the same bin in the mixed testing condition.

We first evaluate the evidence for OI as indicated by a decrease in accuracy as a function of test bin during Test 2. As suggested in the right panel of Figure 4, there is a main
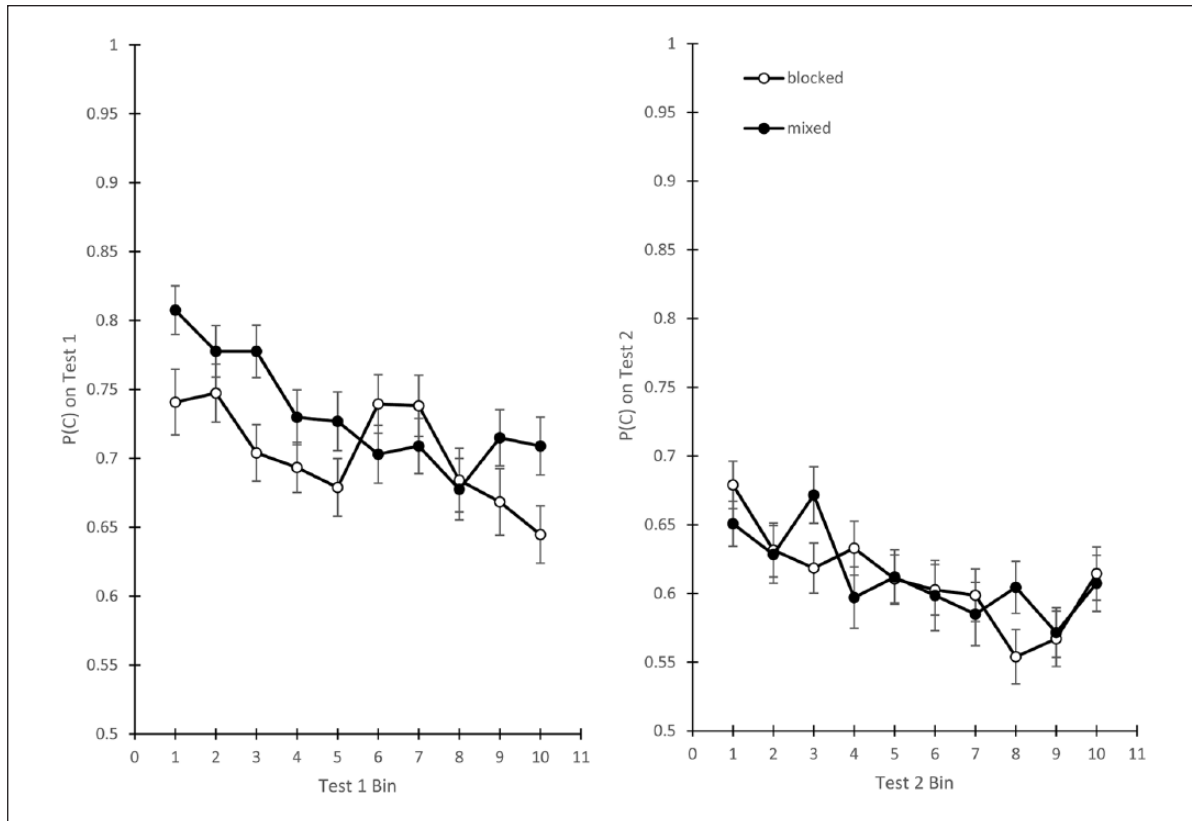
Figure 4. Accuracy from Experiment 2 plotted as a function of test bin. The left panel shows performance on Test 1. Mixed testing shows OI whereas blocked testing shows release from OI at the category switch (between bins 5 and 6). The right panel shows performance on Test 2. Faces and words in Test 2 were randomly mixed and performance is plotted as a function of test type in Test 1. OI was observed in both conditions. Error bars represent ± 1 standard error.

effect of test bin, $F(1, 1269)=4.89$, $p<0.001$, and no interaction with test type (mixed vs blocked; $F(9, 1269)=1.24$, $p=0.27$, BF=121.08); OI is present during Test 2, and the magnitude does not differ as a function of whether Test 1 was mixed or blocked.[2]

We next analyzed accuracy as a function of serial position during Test 1 by dividing Test List 1 into 10 functional "study bins" each with 10 trials.[3] The ANOVA indicates no main effect of study bin, $F(9, 1269)=1.72$, $p=0.081$, BF=10.34, no main effect of the composition of Test List 1, $F(1, 141)=0.02$, $p=0.89$, BF=105.92, and no interaction, $F(9, 1269)=0.97$, $p=0.464$, BF>1000. As evident in Figure 5, recognition of Test 1 foils did not improve following the switch in stimuli, thus providing no evidence that attention is boosted when the stimulus category changes. Further evidence for the lack of attention waning can be found by comparing the bins where release from OI is observed. Accuracy did not differ for Bins 5 and 6 in the blocked conditions, $t(75)=-0.49$, $p=0.628$, BF=7.07. Accuracy did not differ when comparing Bin 6 in mixed to blocked conditions, $t(141)=-0.84$, $p=0.40$, BF=4.03. Thus, we found no evidence supporting the hypothesis that the release from OI results from a boost to attention resulting from the change in stimulus category.
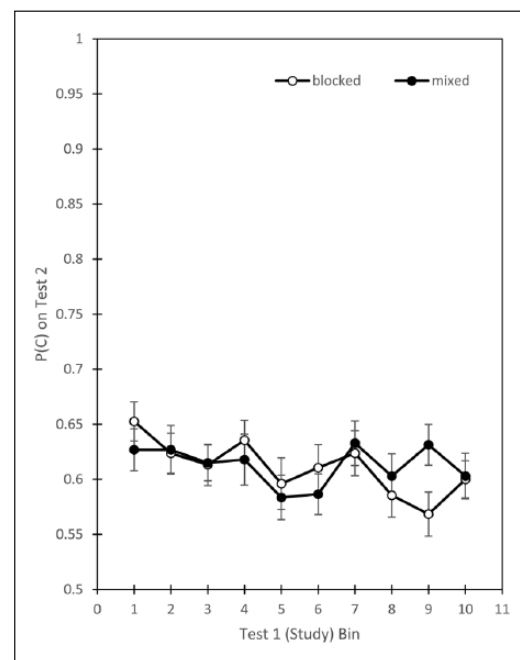

Figure 5. Accuracy on Test 2 of Experiment 2 as a function of Test 1 position. Error bars represent ± 1 standard error.

General discussion

We observed OI for both words and faces. There was a complete release from OI when the test stimuli changed to a different class midway through testing. In addition, the release from OI does not appear to be caused by reorienting to the test upon noticing the change in stimuli because the foil items following the switch were remembered, as well as all other foils. In previous research, we observed an incomplete release from OI when the switch in stimuli midway through a test sequence emphasized a distinction in the semantic features of test stimuli (Malmberg et al., 2012). We replicated and extended that finding and demonstrated that it is not due to enhanced attention from noticing a change.

We propose that OI results from encoding of information during memory testing (Criss et al., 2011; Kilic, 2012; Kilic et al., 2017). Specifically, as implemented, our model assumes that item recognition proceeds as usual with a global match between episodic memory traces and the test item. If an item is recognized as old, the best matching episodic memory trace is retrieved and updated with additional information. When a test item is not recognized but is judged to be new, then a new episodic trace is stored. This results in imperfect storage during the test. Sometimes memory traces are updated with incorrect features (e.g., false alarms) and sometimes a redundant memory trace is stored (e.g., misses). Although other assumptions are reasonable (and perhaps even likely), this simplified model account for the patterns of data observed in the literature (as demonstrated in Criss et al., 2011; Kilic, 2012; Kilic et al., 2017). When non-target traces share features with the retrieval cue used to probe memory, they contribute noise to the information retrieved from memory and cause confusion. The more such traces, the greater the amount of interference that is produced. Across the course of testing, if several items from the same class are presented, interference builds up and memory declines.

The data presented here demonstrate that words and faces are similar within category, resulting in OI but dissimilar across category, resulting in release from OI. The functional division between words and faces is consistent with findings, showing that words and faces do not interfere with one another in single-item recognition (Criss, 2004), associative recognition (Criss & Shiffrin, 2004c, 2005) and cued recall (Aue, Criss, & Fischetti, 2012). Faces and words differ in a number of ways, including different perceptual features of the categories, the verbal nature of words but not faces, the consistent "template" for faces but not words, the evolutionary importance of faces, among others. Isolating the core difference is beyond the scope of this article, what is essential is that words and faces lack cross-category interference. The release from OI for words and faces adds to the finding of Malmberg et al. (2012) that semantic categories of words also produce the same

pattern. In contrast, we find that modality (picture vs audio vs visual presentation) does not result in release from OI (Prince, Criss, Malmberg, & Peckoo, 2013). Together, results of this sort will provide constraints on theories of memory. Like Wickens (1970), we suggest that a full analysis of empirical variables producing release or failure to release from OI is critical to theoretical development and understanding the features and characteristics upon which episodic memory is built.

For instance, context-noise models assume that item information plays no role in recognition memory, rather the prior contexts in which a test item appear are the only source of confusion (Dennis & Humphreys, 2001). Within this modeling framework, OI and release from OI based on item features should not occur. Rather, the context-noise models must make ad hoc assumptions to account for the data such as a decline in the ability to mentally reinstate the study context across the course of testing. The finding observed here—release from OI with words and faces—along with the Malmberg et al. (2012) data provides evidence against this idea, in that the within-class similarity is item-based rather than context-based. One possible solution is to adopt a category cuing mechanism in a context-based model (see Osth & Dennis, 2015 for one example). Of course, item and context noise are not mutually exclusive; in fact, the majority of item-noise models include context-noise. The discussion heretofore emphasized the differences in model approaches; however, a combined approach is not just feasible but almost certainly necessary.

An alternative account is that OI is the result of habituation, and release from OI is the result of a reorienting to the test or restored attention. The idea first arose during the study of PI. For example, Watkins and Watkins (1975) evaluated study order compared to test and found that PI was the result of test, not the result of study. They had participants study trigrams in a PI paradigm. The key manipulation is that they did not test after every trigram and could evaluate performance for successive studied triagrams compared to successively tested triagrams. The critical result for our purposes is that performance on a final test did not change with successively studied triagrams (e.g., the third was remembered as well as the first) but performance did decrease with successively tested triagrams. This and other findings lead Watkins and Watkins to reject the attention hypothesis described earlier in favor of an interference account. They suggest that the buildup of PI stemmed from increasing difficulty in retrieval (e.g., cue overload), and the release caused by changing the item types resulted from unique cues (e.g., Öztekin & McElree, 2007; Watkins & Watkins, 1975).

Likewise, we find no evidence supporting the attention hypothesis for the release from OI observed in recognition testing. We evaluated this by giving a test for foils from the

initial test. If attention is boosted upon noticing the change in stimuli, then performance for the items following the switch should be enhanced. Instead, we found similar levels of accuracy regardless of the position on the initial test suggesting no change in encoding, effort, attention, or reduction in habituation due to the blocked nature of the test. Converging evidence that changes in interference, not attention drive OI come from response time (RT) analysis. RT increases over the course of testing (e.g., Murdock & Anderson, 1975). This could be the result of changes in the quality of evidence, consistent with an interference account or a change in the amount of evidence collected before making a decision, consistent with an attention-based account (or both). Kilic (2012) evaluated these possibilities and found strong evidence for steadily increasing interference and occasional, but inconsistent evidence for a dip in attention at the end of testing. Of course, item interference, context noise, and waning attention are not mutually exclusive mechanisms. To the extent that attention changes during the course of a test, its effects are likely to be in addition to contributions from item or context noise. A fruitful avenue for future research might be to evaluate the relative contributions of each of these mechanisms with both accuracy and RT measures and models.

Collectively, these results anchor the finding that testing helps memory by providing a more nuanced view. Testing episodic memory harms subsequent retrieval attempts of other items. This can be mediated by changes to the material being tested, namely, testing a dissimilar class of stimuli. However, this short-term harm to memory is countered by an improvement to long-term memory for the tested items.

Declaration of conflicting interests

Funding

Notes

1. Note that recall-based tasks of semantic knowledge (e.g., category generation) show slower recall with each successively retrieved item, and this has been interpreted as OI by some (e.g., Roediger, 1974) but not others (Aue et al., 2015; Bousfield & Sedgewick, 1944; Hills, Jones, & Todd, 2012). This debate is beyond the scope of the current article.
2. An exploratory analysis of accuracy on Test 1 on performance on Test 2 was conducted in a 2 (correct or incorrect on Test 1) × 2 (test type—mixed or blocked) × 10 (test bin). Only 83 participants had an observation in each cell for this analysis. Foils that were incorrectly chosen on Test

1 were more likely to be accurately remembered on Test 2 ($F(1, 81) = 51.05$, $p < 0.001$) but this does not change the magnitude of OI (there were no interactions with test type [$F = 0.215$, $p = 0.64$], test bin [$F = 1.15$, $p = 0.33$], and no three-way interaction [$F = 1.25$, $p = 0.26$]).
3. To ensure that the effect was not related to the arbitrarily selected bin size of 10, we plotted accuracy for each trial averaged over participant. There is no indication of any attentional boost on any trial following the change in stimulus type.

References

Annis, J., Malmberg, K. J., Criss, A. H., & Shiffrin, R. M. (2013). Sources of interference in recognition testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1365–1376.

Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal of Memory and Language*, *66*, 109–122.

Aue, W. R., Criss, A. H., & Prince, M. A. (2015). Dynamic memory searches: Selective output interference for the memory of facts. *Psychonomic Bulltein & Review*, *22*, 1798–1806. doi:10.3758/s13423-015-0840-5

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of restricted associative responses. *Journal of General Psychology*, *30*, 149–165.

Criss, A. H. (2005). The representation of single items and associations in episodic memory (Doctoral dissertation, Indiana University, 2004). *Dissertation Abstracts International-B*, *65*, 6882.

Criss, A. H., & Malmberg, K. J. (2008). Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language*, *59*, 331–345. doi:10.1016/j.jml.2008.05.002

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*, 316–326. doi:10.1016/j.jml.2011.02.003

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

Criss, A. H., & Shiffrin, R. M. (2004c). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition*, *32*, 1284–1297.

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1199–1212.

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

Etz, A., & Vandekerckhove, J. (2017). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*. Advance online publication. doi:10.3758/s13423-017-1262-3

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833–848.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, UK: John Wiley.

Grider, R. L., & Malmberg, K. J. (2008). Discriminating between changes in bias and changes in accuracy for recognition memory of emotional stimuli. *Memory & Cognition*, 36, 933–946.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, 119, 431–440. doi:10. 1037/a0027373

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.

Kilic, A. (2012). Output interference and strength based mirror effect in recognition memory. *Psychology – Dissertations*, 174. Retrieved from http://surface.syr.edu/psy_etd/174

Kilic, A., Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2017). Models that allow us to perceive the world more accurately also allow us to remember past events more accurately via differentiation. *Cognitive Psychology*, 92, 65–86.

Koop, G. J., Criss, A. H., & Malmberg, K. J. (2015). The dynamic effects of feedback and test composition during recognition memory testing. *Psychnomic Bulletin & Review*, 22, 509–516. doi:10.3758/s13423-014-0703-5

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

Lockhart, R. S., & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, 74, 100–109. doi:10.1037/h0029536

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer software].

Malmberg, K. J., Criss, A. H., Gangwani, T. H., & Shiffrin, R. M. (2012). Overcoming the negative consequences of interference from recognition memory testing. *Psychological Science*, 23, 115–119.

Malmberg, K. J., Lehman, M., Annis, J., Criss, A. H., & Shiffrin, R. M. (2014). Consequences of testing memory. In B. Ross (Ed.), *Psychology of learning & motivation* (Vol. 61, pp. 285–313). San Diego, CA: Elsevier Academic Press.

Morey, R. (2015). *On verbal categories for the interpretation of Bayes factors*. Retrieved from http://bayesfactor.blogspot.co.uk/2015/01/on-verbal-categories-for-interpretation.html?m=1

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 145–194). Hillsdale, NJ: Lawrence Erlbaum.

Osth, A. F., & Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological Review*, 122, 260–311. doi:10.1037/a0038692o

Öztekin, I., & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language*, 57, 126–149.

Prince, M. A., Criss, A. H., Malmberg, K. J., & Peckoo, S. (2013, November). *The effect of presentation modality on output interference*. Poster Presented at the Annual Meeting of the Psychonomic Society, Toronto, Ontario, Canada.

Roediger, H., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Roediger, H., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, 2, 261–269. doi:10.3758/BF03208993

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

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.

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

Starns, J. J., Chen, T., & Staub, A. (2017). Eye movements in forced-choice recognition: Absolute judgments can preclude relative judgments. *Journal of Memory And Language*, 93, 55–66. doi:10.1016/j.jml.2016.09.001

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60, 158–189. doi:10.1016/j.cogpsych.2009.12.001

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, 1, 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*, 25, 440–445.

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition*, 35, 526–544.