2-4-2015

Understanding Proactive Facilitation In Cued Recall

William Roger Aue
Syracuse University

Recommended Citation

Aue, William Roger, "Understanding Proactive Facilitation In Cued Recall" (2015). *Dissertations - ALL*. 1796.
https://surface.syr.edu/etd/1796

Abstract

Confusion of older information with newer information is a potent source of memory errors. For example, remembering exactly where you parked in a parking garage can be difficult, if it somewhere you frequently park. The reason for the difficulty is because the older memories for parking the garage are easily confused with the most recent one. The current project focused on understanding how memories for recent experiences interact, or interferes, with other related information. In a typical memory interference experiment participants study multiple lists of pairs of items. Items from an initial study list (e.g., A-B) reappear on a second study list paired with new, other items (e.g., A-Br). Performance for A-Br pairs is contrasted with control pairs exclusive to the second study list (e.g., A-B, C-D). In the current series of experiments we used such a paradigm to examine a phenomena called proactive facilitation (PF). This is the observation that the memory for a second presentation of a target (Br) is better when cued by its partner (A) despite being studied with a different partner during its initial presentation. This contrasts proactive interference (PI), a common finding that oftentimes memory is worse in the very same scenario. Indeed a combination of PF and PI appear to be present during recall. When Aue, Criss, and Fischetti (2012) employed such a design they observed PI evidenced by more incorrect responses for A-Br pairs, as well as PF evidenced by more correct responses for A-Br pairs relative to C-D pairs. They proposed multiple explanations for PF and a subset are evaluated in the current series of experiments. I examined three hypotheses in an attempt to understand PF. First, I examined whether it is the case that, in the aforementioned design, participants were more willing to provide a response for A-Br pairs and they simply happen to be outputting both more correct (PF) and incorrect (PI) responses. Second I examined whether participants were spending more time searching memory, resulting in the additional responses being provided. Third, I examined whether participants were encoding the items better the second time they are encountered. In general the data appear to be most consistent with the idea that a portion of items, when encountered a second time, are encoded more completely. Implications for models of memory are discussed.

UNDERSTANDING PROACTIVE FACILITATION IN CUED RECALL

by

William R. Aue

B.A., Butler University, 2004
M.A., University of West Florida, 2008

DISSERTATION
Submitted in partial fulfillment of the requirements for the degree
of Doctor of Philosophy in Experimental Psychology

Syracuse University
December 2014

Acknowledgements

I would like to express my deepest gratitude for my supervisor Dr. Amy Criss without whom this work would not be possible. Over the years she has provided a wealth of guidance for, and patience with, my meanderings into various research topics, not to mention a level of support beyond that which I deserve. I would also like to thank Dr. Marc Howard and Dr. Brad Wyble for providing valuable guidance in the early stages of this project and in my career that has helped shape me into the scientist I am today. A special thanks to Dr. Wyble for agreeing at the last minute to be a part of my defense committee.

I would also like to thank Dr. Asli Kilic, Dr. Pernille Hemmer, and Dr. Greg Koop for being good friends and colleagues. My time at Syracuse would have been far less interesting and my research that much more difficult without their help, suggestions, and friendship.

I would also like to thank my parents and family for their constant support and encouragement. Lastly, I would like to thank my wife, Misty Aue. She has been a source of constant support and tolerance for my frustration venting, late nights spent working, and weekends spent in the lab.

Table of Contents

List of Tables

List of Figures

Introduction

A person’s memory represents a connection to their past. Whether it is memories for a lifetime of experiences (e.g., birthdays, vacations) or a vast amount of acquired knowledge, without access to such information we would be lost in the world. Accordingly, much research has focused on understanding the circumstances under which different memories are acquired or lost. The current project focused on understanding how memories for recent experiences called episodic memory (Tulving, 1985) interacts with other related information that has been previously experienced. For example, there are many studies demonstrating that memories of similar or related information can interfere with one another (e.g., Anderson & Neely, 1996; Postman & Underwood, 1973) making them difficult to retrieve. This is a fairly intuitive result. For example, someone who upon arriving home places their keys in a different spot each day may have difficulty finding where they *last* laid their keys. The problem is that the item used to cue the memory search (i.e., keys) is associated with multiple target locations (e.g., key dish, entryway, kitchen table, coat pocket). The difficulty lies in isolating the most-recent location of the keys in the midst of multiple previous locations. This observation, that memory for previous events interfering with newer ones, has been established experimentally and is called proactive interference (PI).

A task commonly used to measure the impact of such interference on memory is cued recall (CR). Cued recall is a task that tests the ability to retrieve a piece of information when given another related piece of information with which to search memory. For example, retrieving a person’s name when you see their face or retrieving a significant event (e.g., an anniversary) when given the date. The task is particularly useful for understanding memory because this is how people typically use memory in their natural environment, namely by using pieces of information to interrogate memory for other related information.

In the lab, for a typical CR task a participant would study a list of pairs of items (e.g., face-name pairs). Afterward, their memory would be tested by being provided with one member of the pair, such as a face (i.e., the cue), with which to search memory and asked to retrieve the name with which it was recently studied (i.e., the target). We can also manipulate the number of times and the number of pairings in which individual items appears during study. This is the general design employed in the current project and is similar to the traditional paradigm used to understand the impact of interference on memory. A participant might study one list of face-name pairs and then a little later study a second list of face-name pairs where some of the pairs have been rearranged such that a particular face is now presented with a different name. Typically performance is worse for items that are associated with multiple other items such as a face associated with multiple names. If asked to recall the most-recently presented name for a particular face, a participant will do so less often relative to a face that has been paired with only a single name. Recently, however, researchers have observed that under similar circumstances associating a cue (e.g., face) to multiple targets (e.g., names) can also actually facilitate recall, a finding termed proactive facilitation (PF; Aue, Criss, & Fischetti, 2012; Wahlheim & Jacoby, 2013).

The specific purpose of the current project was to understand the mechanism(s) underlying PF in cued recall (CR). I begin with an overview of the research in PI and PF, describing the experimental techniques used and discussing major findings and discrepancies. Then I evaluate three of the potential explanations for the PF that were proposed in Aue, Criss, and Fischetti (2012). I follow these explanations with a presentation of a series of experiments conducted to tease apart some of the explanations of PF.

Table 1: Key terms and abbreviations

Abbreviation	Term	Definition
CR	cued recall	A memory task where, following the study of a list of pairs of items, a participant is given part of the one member of a pair and asked to generate the other item with which it was studied.
MMFR	modified-modified free recall	A memory task where, following the study of multiple lists of pairs of items where items appear on multiple lists (see Table 2), a participant is given one member of a pair and asked to generate all of the other items with which it was studied over the lists.
PF	proactive facilitation	The *improvement* in memory for information associated with conflicting previous experience with the same or similar information.
RF	retroactive facilitation	The *improvement* in memory for earlier information following the recent presentation of conflicting information.
PI	proactive interference	The interference of older, conflicting information for the memory of more recently encountered information.
RI	retroactive interference	The interference of recent, conflicting information for the memory of older, previously encountered information.
RT	reaction time	The measurement of the latency of a given response from a specified starting point.

Proactive interference & facilitation

In the lab, memory interference is sometimes investigated by manipulating the number of associations formed between different items. For example, a participant may initially learn that for a particular cue item (e.g., item A) they should respond with a particular target (e.g., item B) and thus the association A-B is formed. In paired-associate learning, where much of our understanding of associative information developed, a participant repeatedly learned a list of pairs (A-B) until performance for the list reaches a certain criterion (e.g., 90% correct). Next the participant is asked to learn a new list of associations where items from the previous list (e.g., A) are paired with new targets (D) to form an A-D association[1]. In PI participants' learning of A-D pairs is tested by giving participants A and having them recall the most-recent target (D). When compared to pairs without interfering associations, learning of A-D to criterion tends to take longer (e.g., Postman, Stark, & Burns, 1974) and when participants are not required to learn pairs to criterion performance tends to be worse for A-D pairs relative to C-D pairs (e.g., Postman & Gray, 1977). Explanations for PI have included difficulty isolating A-D associations in memory, difficulty learning A-D associations, or response competition at test between B and D, amongst other ideas (Anderson & Neely, 1996; Crowder, 1976; Postman & Underwood, 1973). For additional examples of pair types used in these designs see Table 2.

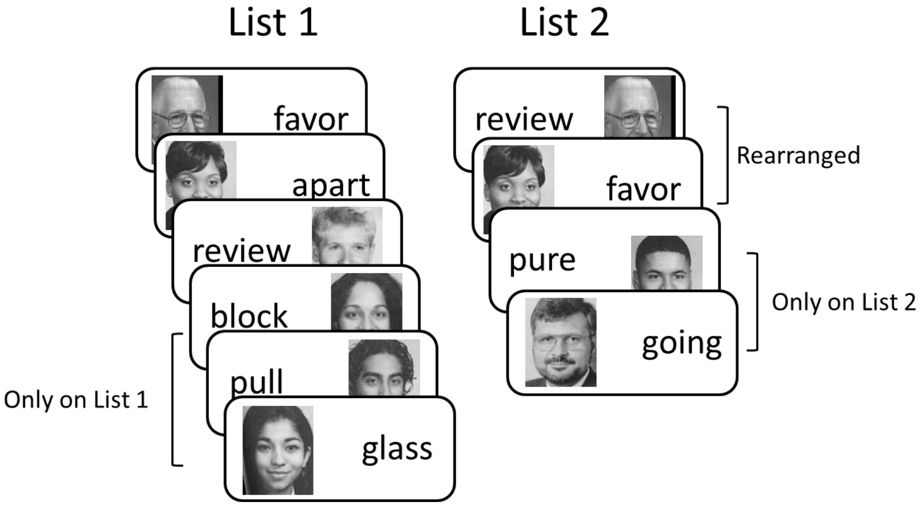

Figure 1. The study design from Aue et al. (2012; Experiment 1). Rearranged pairs correspond to A-Br pairs, List 2 only pairs correspond to C-D pairs. Figure taken from Aue et al. (2012).

---

[1] In the literature, this type of pair is also sometimes referred to as an A-B, A-C pair (e.g., Burton, Lek, & Caplan, 2013; Postman & Stark, 1964; Underwood, 1949).

Table 2. Different types of pairs employed in interference designs

Prior list study		**Potential critical list pairs**			
Cue	Target	Pair type	Cue	Target	Description
Absence	Hollow	A-B, A-B	Absence	Hollow	Both cue and target are repeated together
Pupil	River	A-B, A-B’	Absence	Empty	Repeated cue, semantically similar target
		A-B, A-Br	Absence	River	Both cue and target are repeated but in different pairings
		A-B, A-D	Absence	Tissue	Repeated cue, new target
		A-B, C-B	Pillar	Hollow	New cue, repeated target
		A-B, C-D	Pillar	Tissue	New cue, new target

In contrast to PI, having multiple associations between to-be-remembered cues and targets also sometimes helps memory. Recently, Aue et al. (2012) observed both PI and PF during cued recall using the study design shown in Figure 1. Their approach differed from the PI experiments discussed above in three notable ways. First, unlike paired-associate learning where participants learn pairs to criterion, participants in Aue et al. only experienced the A-B pair once and were not tested on the association. The different technique reflects a focus on examining the nature of associations as opposed to the learning process for associations as is the case for paired-associates tasks. Moreover, others (e.g., Burton, Leik, & Caplan, 2013) have suggested that learning to criterion masks variability in associative interference that underlies PI and PF. The second difference is that both items in some of the second list pairs were studied in the first list. These pairs, called an A-B, A-Br pairs, differs from A-B, A-D pairs where only A (i.e., the cue) had been studied previously. The use of A-Br pair types matches the design employed by Criss and Shiffrin (2005), and has important implications for subsequent explanation of the data, as will be discussed. Third, during the second list both A-Br and C-D pairs appeared in the same study set rather than in separate sets. I included the pairs on the same list to simplify the experimental design, but it is worth noting that it differed from typical PI experiments where only A-D or C-D are tested at a time.

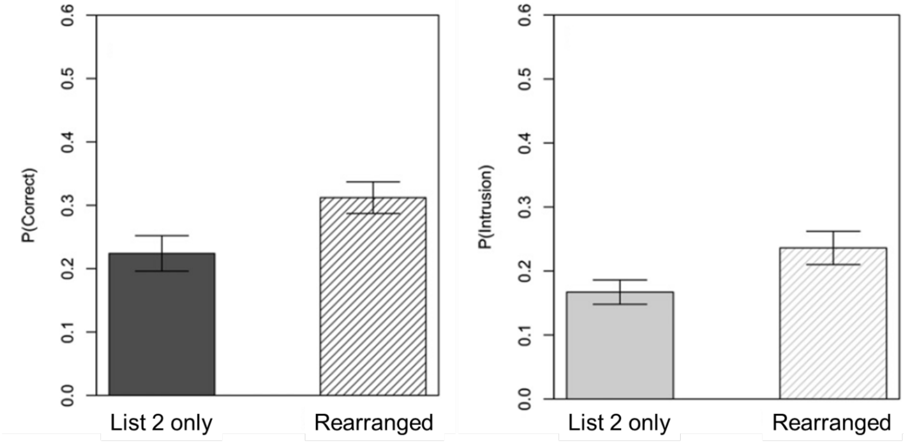

Figure 2. The results of the Aue et al. (2012; Experiment 1). Pairs that were rearranged across the two study lists (i.e., A-B, A-Br) had significantly more correct (left panel) and incorrect (right panel) responses relative to pairs that were exclusive to the second study list (i.e., A-B, C-D).

In the Aue et al. (2012) experiments, participants studied a list of word-face pairs during an initial incidental study list. After a short break participants studied a second list wherein half of the pairs were comprised of items from the first list that had been rearranged into new pairs (i.e., A-B, A-Br, herein referred to as A-Br pairs). The other half of the pairs on the list were items that were only on list 2 (i.e., A-B, C-D, herein referred to as C-D pairs). Participants were then tested on CR where they were shown a face and asked to recall the word it was studied with during the most-recent list. As can be seen in the right panel of Figure 2, when the number of incorrect responses is considered there are more intrusions for A-Br pairs relative to C-D pairs. The increased intrusions for A-Br pairs is evidence of PI. What's more, Aue et al. reported that nearly a third of the intrusions came from the initial B response for A-Br pairs further indicating that the previous B response was interfering with performance. Evidence of PF comes from the correct responses (Figure 2, left panel) where there were significantly more correct responses for A-Br pairs relative to C-D pairs. Thus, participants had more correct and incorrect responses for A-Br pairs relative to C-D pairs, demonstrating both PI and PF. Criss and Shiffrin (2005), which served as the theoretical basis for Aue et al. (2012), observed a similar pattern of data for associative recognition test. Participants were more likely to endorsed A-Br pairs as "old" relative to C-D pairs as evidenced by higher hit rates and higher false alarm rates for rearranged A-Br foils relative to equivalent C-D foils. PF has also been observed using A-B, A-D designs in modified-modified free recall paradigms (MMFR; Barnes & Underwood, 1959). In an MMFR test participants are given a cue and asked to recall *both* responses that were paired with a given cue during study and identify the list in which the response was studied. For example, after studying A-B, A-D, participants are given A and asked to recall both B and D and assign the responses to the appropriate list (e.g., list 1 and 2 respectively). The MMFR technique was developed initially to measure the availability of the two responses in order to understand the nature of associative interference (e.g., Barnes & Underwood, 1959; Tulving & Watkins, 1974). Wahlheim and Jacoby (2013) had participants study two lists where a cue is associated with different target responses on each list (e.g., A-B, A-D) and target responses were unique to their respective lists. They also employed a C-D control condition manipulated within list. At test, Wahlheim & Jacoby (2013) asked participants to recall the D item given a cue (A). Critically, participants were also asked to report if another item came to mind first and if that happened what the items was that came to mind. They found that the recall of D in an A-D pair was greater when participants reported B (from the initial A-B pair) coming to mind *prior* to the recall of D. When participants reported no item coming to mind before D, A-D pairs were recalled worse relative to the control C-D pair. They interpreted these results as being a mixture of proactive interference and facilitation. Likewise, Burton et al. (2013), employing a similar design, observed a positive correlation between recall of B and D given A as a cue. Postman and Gray (1977) also observed that when given A while having studied A-B, A-D, participants were most successful at retrieving D if they first recalled B. These results, in different experimental designs provide corroborating evidence of the reliability of PF. Next we turn to potential explanations for PF as observed in the current design.

Proposed explanations for PF in Aue et al. (2012)

Aue et al. (2012) proposed a number of explanations to account for the results in cued recall. These proposals were couched in the conceptual framework of the retrieving effectively from memory (REM; Shiffrin & Steyvers, 1997) model, and drew on the recall versions of the model in particular (e.g., Diller, Nobel, & Shiffrin, 2001; Malmberg & Shiffrin, 2005). However, as discussed in Aue et al. (2012) the current versions of REM is insufficient to account for the cued recall data due to lacking associative information beyond the simple co-occurrence of item information. As such we were not necessarily confined to any particular explanation or framework. I have considered three potential explanations for the data in Aue et al.: 1) a change in response threshold, 2) better encoding of list 2 pairs, or 3) longer memory searchers in response to familiar cues (e.g., A). First, I'll begin with a brief overview of each explanation before presenting the experiments testing these ideas.

Differences in response threshold.

Given the pattern of data that we observed in Aue et al. (2012; Figure 2) wherein more responses (both correct and incorrect) were provided for A-Br pairs relative to C-D pairs one potential explanation is that participants are changing the quality of responses they are providing. In the Diller et al.'s (2001) cued recall part of the REM model this would manifest as a change in the quality of the cue that is considered acceptable for sampling. Herein I'll refer to this as the sampling threshold. At test, a face is provided and it is matched to the other faces in memory. The faces in memory are sampled probabilistically based on how well they match. Once sampled if the similarity between the sampled trace and the face cue does not pass a specified threshold then the search is begun anew or terminated and no attempt at recovering the target (i.e., the word) is made. For the Aue et al. data, participants could be more (or less) willing to accept a sampled memory trace and attempt recovery for either the A-Br pairs or C-D pairs. A contrived real-world example of this would be choosing who to potentially take on a date from a match-making website based on a compatibility score. If the compatibility score for a sampled individual is greater than some threshold (e.g., 75% match) then the person may be selected for a date. The threshold can be changed based on circumstances. For example, if a person is particularly desperate then they may set a liberal criterion accepting anyone above 50% match. Likewise, they can also set a stricter criterion such as only accepting people who match at 90% or better. In this same way, participants could be changing the quality of the sampled trace that they are willing to accept for A-Br pairs or C-D pairs. However, this threshold would not account for the current data given that it would need to be based on a cue that has yet to be sampled. If it were the sampling threshold then, a list-wise threshold change would be most likely which would influence both A-Br and C-D pairs equally.

A second possibility, although not part of the Diller et al. (2001) model, is a change in the response threshold during the recovery of the target that is associated to a sampled cue. Diller et al. assumes that if the target is recoverable (because enough information is stored) then it is output. However, as will be discussed below, others have suggested that we also evaluate the quality of a recovered response prior to output. Changes in this response threshold would manifest as participants being more (or less) willing to output a response for a given cue. Returning to the dating site example, the compatibility score says nothing about whether a person is relationship material. For instance, the decision to pursue a relationship could depend on the quality of a first date. If it went well and both persons enjoyed themselves, then they may be more apt to pursue a relationship versus if it went poorly. Again factors such as desperation or quality of alternatives can change the threshold of date quality needed to decide to pursue a relationship. In Aue et al. (2012) participants had the freedom to decide whether or not to respond to a given cue. It is possible that a participant may have retrieved a response but chose to withhold the response because they were unsure about its quality. If participants systematically changed their willingness to respond in one condition relative to another it would be evidence of a shift in their response threshold.

Koriat and Goldsmith (1994, 1996) discussed the role of the response threshold as it relates to how people decide whether to report items retrieved from memory. Koriat and Goldsmith (1996) provided a schematic model of memory monitoring. In the model, once responses are retrieved they are compared to a response threshold to decide whether they are good enough to report. If the quality (e.g., confidence in the response) of the response exceeds the response threshold, the response is output otherwise it is withheld and no response is provided. Koriat and Goldsmith further demonstrated that a participant's willingness to provide responses (i.e., response threshold) can be influenced experimentally by manipulating incentive structures in the experiment such as rewarding correct responses or penalizing incorrect responses. For example, when the penalty for providing an incorrect response is high participants tended to adopt a stricter response threshold as evidence by fewer but highly accurate responses (Koriat & Goldsmith, 1994). In contrast, setting a less strict penalty resulted in more responses overall that tended to be less accurate. Moreover, the influence of response threshold is negated when participants are required to provide a response to every cue, a task called forced recall. This is because participants simply output the response regardless of where it falls with respect to response threshold (Koriat & Goldsmith, 1996).

The Aue et al. (2012) data could be partially explained by either a liberal or conservative response threshold. A liberal response threshold would manifest as participants providing responses that they would normally withhold because it is of low quality (i.e., low memory strength, confidence). The result would be more responses provided overall and would drive up both correct and incorrect responses, as is the case for the A-Br pairs. Alternatively, participants could be adopting a conservative threshold where they are withholding responses that they otherwise would have provided. In this case fewer responses meet the response threshold so overall responding is reduced for both correct and incorrect responses, as is the case for C-D pairs.

Such a shift in response threshold could be driven by the familiarity of the cue at test. Recall that participants have seen the individual items in A-Br pairs in both the first and the second list, but in different pairs. At test, participants are provided with a cue and asked to recall the word that it was studied with during the most-recent list. Thus, the strength of cue is the only viable[2] basis for setting the response threshold on a trial-by-trial basis. The decision to adopt a liberal response threshold for A-Br pairs could be based on the familiarity of the cue. Cue familiarity is higher for individual items in A-Br pairs than in CD pairs as evidenced by higher hit rates in tests of single-item recognition (Aue et al., 2012; Criss & Shiffrin, 2005) and increased accuracy in free recall (Hirshman, Burns, & Kuo, 1993). Additionally, this is consistent with metacognitive research finding that the cue familiarity can influence subjective feeling of knowing of a target response (e.g., Schwartz & Metcalfe, 1992; Metcalfe, Schwartz, & Joachim, 1993), retrieval strategy (e.g., Reder, 1987), and willingness to provide a response (e.g., Hanczakowski, Pasek, Zawadzka, & Mazzoni, 2013). Likewise, participants could adopt a conservative response threshold for C-D pairs due to the lack of familiarity of C as a cue relative to A-Br cues.

In the Experiment 1 and 2 I investigated whether I could replicate the pattern of data observed in Aue et al. (2012) employing the same general design as Aue et al. (2012) but changing the testing scenario to try to differentiate between these alternatives for response threshold.

Better encoding for A-Br pairs during list 2.

Another possible explanation for PF discussed in Aue et al. (2012) is a possibility that list 2 pairs were better encoded. For example, during the study of list 2 participants could be storing more accurate versions of the A-Br pairs relative to C-D pairs. This could take the form of

---

[2] It bears noting that test context is also available and presumably used at test. We are presuming here that the context information is constant for A-Br and C-D pairs given that they were experienced during the same list and have a similar same study-test lag.

stronger item information (i.e., memories for the individual items), stronger associative information (i.e., information indicating that two items were studied *together*), or both. Conceivably, this could stem from the participants prior familiarity with items in A-Br pairs, having just studied them in list 1 in different pairs. One possible way this could work is if when the A-Br items appear on list 2, the participants draw on the existing representation of the item in episodic memory to store updated versions of those traces (i.e., strengthen it) and add them to the representation for the new pair rather than storing only a new representation for the List 2 pair.

Hirshman, Burns, and Kuo (1993) have posited a similar updating mechanism for a type of proactive facilitation that they have observed for free recall. They observed that when participants study A-B, A-D and are asked to recall as many of the items from the second list (in this case A-D) as possible, they recall more items relative to the when participants study A-B, C-D and are tested on C-D. Thus, they observed better performance for individual items in list 2 when part of the pair is repeated across lists which is consistent with subsequent research (e.g., Aue et al., 2012; Criss & Shiffrin, 2005). Their explanation for the result assumes that pairs are stored as separate items and an association of the items. When participants study A-B followed later by A-D, they simply update the existing memory for A and associate it to the new trace for D. As a result, there are only three items in memory (i.e., A, B, & D). The facilitation for free recall of D is the result of A-D pairs having a shorter "'list" relative to C-D pairs where two new memory traces are stored (i.e., one each for A, B, C, D). Thus, this represents a different perspective of what could constitute better encoding. In this case, the existing information for A is updated to contain the new information associating it to D, although this explanation would not necessarily explain facilitation in CR. Another related idea is that when an item is encountered a second time and it is identified as having been seen before a recursive representation is stored that containing information about both the initial and subsequent presentations of the item (Hintzman, 2004). This explanation has been applied to PF (Wahlheim & Jacoby, 2013) suggesting that when A-B, A-D are studied they are stored in a single trace containing the item and order information (e.g., A-B-D). At test, the recall of B given A facilitates the correct recall of D. I give this idea a more thorough discussion later in the general discussion.

It is also important to note that a better encoding explanation alone would seem an inadequate explanation of the data. If it simply were better encoding then we might expect just to observe PF in the absence of PI. If better encoding of list 2 pairs is found to be a factor it will likely be a combination of factors (e.g., better encoding and longer searches) that explain the data.

In Experiments 3-6, I investigated the question of whether list 2 pairs were better encoded using two general approaches. First, I added pairs where either just the cue or just the target had been studied previously. These correspond to A-B, A-D pairs and A-B, C-B pairs from Table 2, respectively. They were included in the same study list as the A-Br pairs and C-D pairs. This design has the potential to address at least two relevant questions. First, the A-D and C-B pairs are mixed-repetition pairs given that only one item in the pairs has been studied before. Hereafter I equate repetition to strength and as such refer to repeated items as ‘strong’ and non-repeated items as ‘weak.’ If better versions of pairs are being stored when both items have been experienced in list 1, then it is reasonable to expect that performance for mixed-strength (e.g., A-D and C-B) pairs would be lower than pairs where both (A-Br) were repeated and higher than when neither are repeated (C-D). Second, this design has can also inform which component (i.e., cue, target) of the to-be-remembered pair is driving the change in performance observed by Aue et al. (2012). I have suggested that a change in response threshold or longer search explanations would be driven by the familiarity of the cue. I explicitly tested that assumption by manipulating the strength of the cue and target orthogonally. I further examine this idea in Experiment 4 where list 1 was comprised of individual items, instead of pairs, and A-Br pairs are comprised of a repeated cue, a repeated target, or both.

In Experiment 5, I examined performance for list 1 pairs in a retroactive interference design. The design is the same as shown in Figure 1 except at test participants are asked to recall the associations from initial study list (i.e., list 1; A-B pairs). If the benefit for A-Br pairs is occurring during the encoding of list 2, then the representation for list 1 should remain unchanged. The goal was to test whether retroactive facilitation is present for A-Br pairs across lists.

Lastly, in Experiment 6 I manipulated the delay between list 2 study and test in order to manipulate the familiarity of a the cue at test. The aim was to encourage participants to reinstate the list 2 context at test rather than relying on the context available at test to search memory. If the A-Br advantage data are driven by some aspect of cue familiarity then A-Br performance should be similar to C-D performance when the memory search is isolated to list 2 given that cues for both appear only once during list 2. However, if participants are encoding A-Br pairs better than C-D pairs then the advantage should persist over the delay.

Longer Memory Searches.

The third explanation that I consider is that participants are spending more time searching memory for A-Br pairs relative to C-D pairs. In recall models of REM (e.g., Diller et al., 2001; Malmberg & Shiffrin, 2005) search length is represented by a counter index (K) of the number of unsuccessful retrieval attempts that are allowed before deciding that the answer is not known

(K_{max}). For example, during a search a memory trace is sampled probabilistically based on how well it matches the cue provided at test. If a trace is selected and it is strong enough to pass the sampling threshold an attempt at recovering the target from the trace occurs and is either successful or unsuccessful. If the target recovery is unsuccessful, or if no trace is sampled, then the index (K) is incremented. The process repeats until a specified number of search attempts (K_{max}) has been reached. In everyday life, this would be similar to encountering a person that seems familiar but their name escapes you. The number of search attempts is akin to the amount of time you might spend trying to figure out who the person is or from where you know them.

In the Aue et al. (2012) data, both PF and PI could potentially be explained by the fact that participants are spending more time searching for A-Br pairs than C-D pairs. In fact, Diller et al. (2001) proposed that K_{max} is determined by the familiarity of the cue. In Aue et al. with a given cue (A) from an A-Br pair they may sample the target trace from presentation from the first list (i.e., A-B) or the second list (i.e., A-Br). If it is the first list and they recover the target, then it will be the list 1 partner and an intrusion. If it is the second list and they can recover the target, it would be correct. As with the response threshold explanation the process would likely be driven by the familiarity of the cue given that it is the only information provided during a CR test and would be consistent with Diller et al. If participants are spending more time searching for A-Br pairs relative to C-D pairs, then this should manifest as longer reaction times (RT) for A-Br pairs relative to C-D pairs. I address more specific predictions in the introduction to Experiment 7. I tested this idea in Experiment 7 by measuring response time during CR using a modified experimental design.

Tests of Differences in Response Threshold

As discussed above, one possibility for explaining PF is that participants are adopting a different response threshold for either the A-Br pairs or the C-D pairs. In effect, participants could be changing the quality of response that they are willing to output during a test. The A-Br advantage observed by Aue et al. (2012) could manifest in two ways: 1) participants could be providing responses to an A-Br cue that they otherwise would not (i.e., adopting a liberal response threshold), 2) participant could be withholding responses to a C-D cue that they otherwise would provide (i.e., adopting a conservative response threshold). In both cases it could be familiarity of the cue – or lack thereof for C-D pairs – that is driving the effect. In the following two experiments I test these ideas using two approaches. First, I employ instructional manipulations to encourage participants to set a higher response threshold (Experiment 2) or to eliminate the influence of response threshold (Experiment 1). With respect to the aforementioned response threshold explanations, if it the case that participants are withholding C-D responses they would otherwise provide then forcing participants to response (Experiment 1) should selectively boost C-D responses. If, however, participants are providing A-Br responses that they would have otherwise withheld then asking participant to adopt a more strict response threshold (Experiment 2) should attenuate the effect.

As a second metric for evaluating the quality of response being provided for Experiments 1 and 2, I am also collecting retrospective confidence ratings. Following each response entered, participants are asked to report how confident they are that the response they just provided is correct. If participants are changing the quality of response they are providing for either A-Br or C-D pairs, then it may be reflected in their confidence ratings. If participants are providing lower quality responses for A-Br pairs that would have otherwise been withheld then they would be associated with a lower confidence rating relative to the C-D pairs. The same pattern would be observed if participants were withholding responses to C-D pairs.

To summarize, if the difference in response threshold is present, then performance should be influenced by the instructional manipulations in Experiments 1 and 2 and reflected in the retrospective confidence ratings provided by participants.

Experiment 1: Forced-Report Cued Recall

Of the two possibilities that were outlined for a response threshold explanation, I'll first address the issue of a conservative response threshold for C-D pairs. Perhaps it is the case not that A-Br pairs are advantaged but rather that responding to C-D pairs is attenuated in response to the less familiar cue, relative to the familiarity of the A-Br cues. To test this, I adopted a forced report procedure in the first experiment. The testing scenario is the same as in Aue et al. (2012); however, participants are required to provide a response to each cue before proceeding. The rationale is that if it is the case that participants are withholding C-D responses, then forcing them to respond will result them in outputting the best matching target regardless of where the response threshold is placed effectively eliminating the response threshold in this experimental design. If participants are setting a more conservative response threshold for C-D pairs and this threshold is driving the apparent advantage for A-Br pairs, then performance for the two conditions should be equated under forced report.

Method.

Participants. Forty participants from the Syracuse University subject pool participated in the experiment for course credit. All participants were included in the analyses, and the sample size was determined by approximating that of Aue et al. (2012). The Syracuse University

Institutional Review Board approved all study protocols, and all participants provided written informed consent. This is true of all subsequent experiments unless specified otherwise.

Materials. All of the discussed experiments involve the study of word-face pairs. The face stimuli used are the same as those described in Aue et al. (2012) and Criss and Shiffrin (2005). There were a total of 210 faces used that were standardized for orientation position of facial landmarks. The words used in the experiments were 800 high-frequency words ($M =$ 130.66 from Kučera & Francis [1967] or, alternatively, log frequency $M = 10.46$ in the Hyperspace Analog to Language corpus of Balota et al. [2007]; Lund & Burgess [1996]). These materials were used in all experiments and remained consistent across experiments.

Design. All experiments followed the same general design and procedure that is described in this section, with the critical changes described in each experiment's respective section. Participants completed the experiment individually on a Windows-based computer running Authorware (v. 7.0). The experiment was a within-subject design with two conditions (i.e., A-Br vs. C-D pairs). Participants studied the two lists of word-face pairs followed by a cued recall test. The composition of the critical list (list 2) is described next. The experiment lasted approximately 30 min.

List 1 and list 2 each contained 24 word-face pairs. List 2 was comprised of 12 unique pairs (C-D pairs) plus 12 pairs of items rearranged from list 1 (A-Br pairs). The pairing of A-B and C-D pairs and the re-pairing for A-Br pairs was randomly chosen for each participant. Pairs were presented with the items side-by-side for 3 s, with the left/right position of the face and word randomly chosen on each trial for each participant. Following the presentation of a pair, participants performed an encoding task. Different encoding tasks were used for list 1 and list 2. The tasks are the same as those described in Aue et al. (2012) and Criss and Shiffrin (2005). No differences have been observed for cued recall performance when encoding tasks were switched (Criss, Aue, & Smith, 2011). For the list 1 encoding task (A-B pairs), participants were asked to rate the degree of association between the items on a 9-point scale (1 – *not at all associated* to 9 – *highly associated*). Upon completion of list 1, participants read a comic for 60 s. For the list 2 encoding task (A-Br & C-D pairs), participants were asked to generate a sentence about each pair and then rate how difficult it was to do so using a 9-point scale (1 – *very easy* to 9 – *very difficult*). After the second study list, participants completed a distractor task for 60 s where they kept a running summation of 20 individually presented digits.

The cued recall test immediately followed the distractor task. During test, faces from list 2 appeared one at a time. Participants were asked to type the word that the presented face was paired with on the *most-recent* list (i.e., list 2). The order of the faces was randomized anew for each participant, and the test was self-paced. For Experiment 1, participants were instructed to provide a response to each face cue, even if doing so required guessing. After each recall attempt, participants were asked to report their confidence that the word they just recalled was accurate. Reporting was done on a 6-point scale (*1 – I am sure it is wrong* to *6 – I am sure it is correct*). Participant's responses were coded as either correct or incorrect. A correct response consisted of a response that was either the target word or contained minor errors such as misspellings (e.g., braec instead of brace) or added suffixes (e.g., walked instead of walk). All other responses were coded as intrusions. Occasionally a participant would disobey the instructions and enter a response that indicated they did not recall the cue for a given target (e.g., "no", "idk"). These were coded as No Recall responses and were infrequent[3].

---

[3] No recall responses occurred on M = .002 of trials for A-Br pairs and M = .004 of trials for C-D pairs.

Given that the experiments are designed to induce interference there are many incorrect responses provided during testing. Aue et al. (2012) examined incorrect responses for meaningful patterns, and we do the same here. To that end we classified incorrect responses into one of five categories. For A-Br pairs, if the response provided was the partner of the cue during the first study list then it was classified as a "List 1 partner response." If the response provided was a studied word that appeared in *both* List 1 and List 2 it was classified as a "List 1 and List 2 response." If the response was a studied word that appeared only in List 1 or only in List 2, it was classified as such. Lastly, if the response was not a studied word it was classified as a random word error.

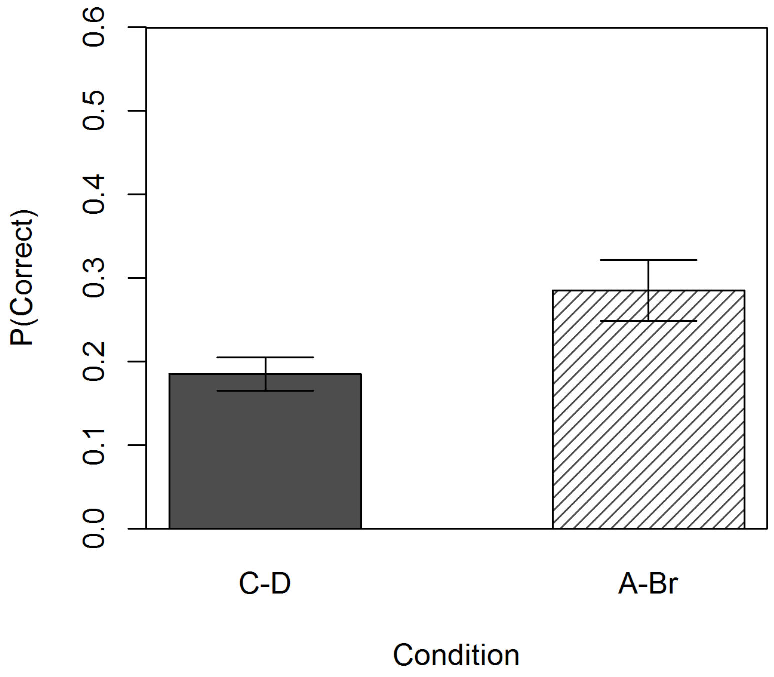

Figure 3. Correct responses as a function of pair condition. Participants had more correct responses for A-Br pairs relative to C-D pairs. Thus forcing participants to respond did not improve performance for C-D pairs indicating that a conservative response bias is not at play.

Unless otherwise specified, data was analyzed using the R programming language (R Core team, 2013) and the "EZ" package (Lawrence, 2013) for statistical analysis.

Results & discussion. If a conservative response threshold is being employed for C-D pairs causing participants to withhold responses relative to A-Br pairs then forcing participants to respond should boost performance for C-D pairs by encouraging them to output the responses that are otherwise being withheld.

Memory performance. To analyze CR performance, I performed a one-way ANOVA on the proportion of correct responses provided for the A-Br versus C-D conditions. Participants provided significantly more correct responses for A-Br pairs ($M = .285$, $SE = .036$) relative to C-D pairs ($M = .185$, $SE = .02$; $F(1, 39) = 12.73$, $p < .001$, $\eta_G^2 = .069$) demonstrating the same pattern as the Aue et al. (2012) data. As it relates to the predictions made by a conservative response bias account, forcing participants to respond did not improve performance for C-D pairs relative to A-Br pairs. This observation is consistent with experiments investigating memory performance under forced versus free recall (e.g., Roediger & Payne, 1985). Specifically, in the absence of an instructional response bias manipulation forcing participants to respond only tends to increase the number of incorrect responses. This indicates that participants having good memory calibration (e.g., Koriat & Goldsmith, 1996) for what they know and do not know when they are free to not respond. For completeness, however, in the next section we examine participant's retrospective confidence ratings for recalled items to be sure that the quality of correct responses provided did not differ across conditions.

In the current experiment, incorrect responses were essentially the inverse of the correct responses, so they are not particularly informative. However, we can examine the type of incorrect response that participants made and compare that relative to the number of intrusions that were possible for each condition based on the number of words studied in each category. This analysis is identical to the analysis performed on intrusions in Aue et al. (2012). Expected value proportions are drawn from the number of intrusions possible for a given category. For example, words that were studied in both list 1 and list 2 were a common intrusion, however only 12 words comprise this set for C-D pairs during an experimental session. This differs for A-Br pairs where set is 10 words here one word is the target and another is the partner of the cue during list 1 study. These intrusion types are outlined further in Table 3. I tested whether the number of responses from each category differed from what would be expected based on the number of intrusions possible for each condition using a Chi-squared test, although it bears noting that given that the analysis violates the assumption of independence. For A-Br pairs, the intrusions differed significantly from the expected values derived from the number of intrusions possible for each category (χ^2 (3, N = 182) = 356.55, $p < .001$). The observed intrusion proportions for C-D pairs also differed significantly from the expected values (χ^2 (2, N = 191) = 78.28, $p < .001$). A similar pattern was observed for Aue et al. and suggests that intrusions are being driven by a process other than response availability.

Table 3: Frequency of certain incorrect response types for Experiment 1

	A-Br			C-D		
	Possible	Expected	Observed*	Possible	Expected	Observed*
List 1 only	12	62.4	18	12	65.49	17
List 2 only	12	62.4	41	11	60.02	56
List 1 partner	1	5.2	45	0	NA	
List 1 and List 2	10	52	78	12	65.49	118

Note. * $p < .05$

Memory confidence. I examined confidence ratings across conditions by averaging responses within subject separately for correct and incorrect responses. If participants are withholding responses for C-D pairs by setting a higher response threshold one might expect participants to be more confident in the responses provided. Consistent with recall performance, there was no difference across conditions for confidence in correct responses. Participants were as confident in their correct responses to A-Br pairs (*M* = 4.61, *SE* = .186) as they were for C-D pairs (*M* = 4.33, *SE* = .1; *F*(1, 68) = 1.03, p = .312, η_G^2 = .015). However, participants tended to be more confident in their intrusions for A-Br pairs (*M* = 2.23, *SE* = .130) than for C-D pairs (*M* = 1.83, *SE* = .092, *F*(1, 68) = 6.13, p = .015, η_G^2 = .073), which is to be expected given the interference design. Moreover, participants were more confident in the correct responses (*M* = 4.55, *SE* = .144) relative to incorrect responses (*M* = 1.98, *SE* = .097; *F*(2, 76) = 115.3, p < .001, η_G^2 = .752) suggesting their responses were well calibrated and that participants would have likely withheld the low confidence responses if they had the option.

Lastly, I examined memory confidence across the classes of incorrect responses. As one might expect, when participants responded with the “List 1 partner” for A-Br pairs, they tended to be more confident in that response relative to the other types of intrusions. The higher confidence for “List 1 partners” was confirmed by a two sample *t*-test comparing “List 1 partner” to “List 1 and List 2 responses” since both responses were studied twice (*t*(73.20) = 5.71, p < .001). Predictably, participants appeared to be least confident in their responses that were unstudied words. No clear pattern emerges from the confidence ratings for the C-D pairs.

Given that the relationship between A-Br and C-D pairs did not change when participants were forced to respond, this suggests that participants are not withholding responses for C-D pairs. Moreover, the fact that I observed no differences in mean memory confidence for correct responses suggests that similar quality responses were being provided for both types of pairs. This indicates that a conservative response threshold for C-D pairs is not a likely explanation for the PF observed by Aue et al. (2012). However, these data do not speak to whether participants are setting a more liberal response threshold for A-Br pairs when they are free to respond. I address this question next.

To summarize, I anticipated that if participants were adopting a conservative response threshold for C-D pairs that resulted in fewer responses being offered, then forcing participants to respond should selectively improve performance for C-D pairs. However, that was not the case in the current data. Instead, the A-Br advantage persisted. Moreover, there were no differences in retrospective confidence ratings to correct responses to indicate a different in memory quality between A-Br and C-D pairs. Next we examine whether a liberal response threshold for A-Br pairs may be resulting in the advantage.

Table 4: Incorrect response confidence by pair condition for Experiment 1

	A-Br			C-D		
	N	*M*	*SD*	N	*M*	*SD*
List 1 only	9	2.45	1.31	11	2.63	1.52
List 2 only	23	2.18	1.22	27	2.05	1.20
List 1 partner	27	3.40	1.30		NA	
List 1 and List 2	33	2.16	.977	35	2.06	.848
Unstudied	32	1.83	.815	36	1.52	.713

Note. N represents the number of participants committing a given type of incorrect response.

Experiment 2: High Confidence Responses

Whereas in Experiment 1 I induced the most liberal response threshold by requiring a response on every trial, here I do the opposite. The aim of the current experiment was to induce participants to adopt a conservative response threshold by asking participants to restrict responses to high confidence responses. If it is the case in that in Aue et al. (2012) participants were setting a more lenient response threshold for the A-Br pairs relative to the C-D pairs. Such a threshold would likely result in increased response rates. As such, asking participants to restrict their responses to only those responses that they were highly confident in should attenuate responses for A-Br pairs.

Method.

Participants. In total 40 Syracuse University undergraduates participated in the experiment. Participants received course credit for their participation.

Materials. The materials were identical to previous experiments.

Design. The details were the same as those described in Experiment 1 with a minor modification to the testing procedure. At test participants were provided with a cue and asked to recall the target it was studied with on the most-recent list (i.e., list 2). Additionally, they were instructed only to respond if they felt highly confident that they were correct, and if they did not feel a high level of confidence they should indicate that they did not know the answer (i.e., withhold their response). Afterward, they rated the confidence in their response using the same procedure as Experiment 1. All other aspects of the experiment were identical to Experiment 1.

Results & Discussion. First, I wanted to be sure that the manipulation of asking participants to respond with higher quality responses was effective relative to the forced report condition in Experiment 1. I took two approaches to demonstrate this. First, I examined whether fewer correct responses were provided in Experiment 2 relative to a "neutral" experiment where the design is identical but where the response threshold was not influenced experimentally (i.e., Aue et al. (2012; Experiment 1). Indeed participants provided fewer correct responses across condition in Experiment 2 of the current paper (*M* = .202, *SE* = .016) when asked to adopt a high threshold relative to Aue et al. (*M* = .267, *SE* = .019; *t*(138.24) = 2.62, *p* = .009, *d* = .433)[4]. Next, since I do not have confidence ratings for Aue et al. I compared average confidence of correct responses collapsed across conditions in Experiment 1 and Experiment 2 of the current paper. If the manipulation were effective, participants in Experiment 2 should be more confident in their responses overall than participants in Experiment 1. Indeed participants in Experiment 2 tended to be more confident in their correct responses (*M* = 5.04, *SE* = .083) than participants in Experiment 1 (*M* = 4.63, *SE* = .098; *t*(414.43) = 3.19, *p* = .002, *d* = .306). Therefore, the manipulation was effective at eliciting higher quality responses. Next I turn to whether asking participants to restrict their responses had the intended impact on performance.

Memory performance. A one-way ANOVA revealed that participants tended to provide a greater proportion of correct responses to A-Br pairs (*M* = .229, *SE* = .024) relative to C-D pairs (*M* = .175, *SE* = .02; *F*(1,39) = 3.89, *p* = .055, η_G^2 = .037). This pattern was borderline significant, however given the fact that it replicates the pattern Aue et al. (2012) observed I am inclined to trust the data are reliable. The same pattern was observed for incorrect responses to A-Br pairs

---

[4] Additionally, an astute reader will notice that correct performance for both A-Br and C-D pairs is also poorer in Experiment 1 relative to the corresponding pairs in Aue et al. (2012; Experiment 1). This likely can be attributed to the additional output interference (Criss, Malmberg, & Shiffrin, 2011; Malmberg, Criss, Gangwani, & Shiffrin, 2012) caused by forcing participants to respond to each cue.

(*M* = .298, *SE* = .026) relative to C-D pairs (*M* = .229, *SE* = .028; *F*(1,39) = 6.40, *p* = .016, η_G^2 = .04), replicating the pattern observed in Aue et al. (2012).

For incorrect responses, the same chi-squared test as in Experiment 1 was performed on the number of intrusions. The intrusions for A-Br pairs (χ^2 (3, N = 118) = 653.91, *p* < .001) differed significantly from the expected values based on response availability. The same was true for C-D pairs as well (χ^2 (2, N = 89) = 21.29, *p* < .001). The data are summarized in Table 5.

Memory confidence. The confidence data for Experiment 2 was similar to the data from Experiment 1. A one-way ANOVA on confidence ratings for correct responses revealed that participants did not differ in their confidence of correct responses for A-Br pairs (*M* = 4.94, *SE* = .144) relative to C-D pairs (*M* = 5.10, *SE* = .177; *F*(1,68) = .473, *p* = 493, η_G^2 = .007). Again, participants tended to be more confident in their intrusions to A-Br pairs (*M* = 3.91, *SE* = .169) relative to C-D pairs (*M* = 3.26, *SE* = .204; *F*(1,72) = 6.13, *p* = .015, η_G^2 = .078). Having participants set a stricter threshold for responding did not influence PF. Thus, there is no evidence that participants are setting a more liberal criterion for A-Br pairs relative to C-D pairs.

Table 5. Frequency of incorrect response types for Experiment 2

	A-Br		C-D	
	Expected	Observed*	Expected	Observed*
List 1 only	40.46	12	27.98	13
List 2 only	40.46	16	30.51	27
List 1 partner	3.37	49	NA	
List 1 and List 2	33.71	41	30.51	49

Note. * *p* < .05; Expected values are based on the same number of possible responses provided for Experiment 1.

Lastly, I examined memory confidence across the classes of incorrect responses (Table 6). As one might expect, when participants responded with the "List 1 partner" for A-Br pairs, they tended to be more confident in that response relative to the other types of intrusions. Again, the higher confidence for "List 1 partners" was confirmed by a two sample *t*-test comparing "List 1 partner" to "List 1 and List 2 responses" since both responses were studied twice ($t(86.52) = 2.57$, $p = .012$). Predictably, participants appeared to be least confident in their responses that were unstudied words. No clear pattern emerges from the confidence ratings for the C-D pairs.

Table 6. Incorrect response confidence by pair condition for Experiment 2

	A-Br			C-D		
	N	*M*	*SD*	N	*M*	*SD*
List 1 only	9	3.56	1.33	10	3.45	1.38
List 2 only	12	3.63	1.03	19	3.38	1.44
List 1 partner	26	4.32	1.38		NA	
List 1 and List 2	25	3.67	1.22	22	3.40	1.34
Unstudied	13	3.49	1.16	17	3.20	1.55

Note. N represents the number of participants committing a given type of incorrect response.

Response Threshold Discussion

The data from the first two experiments indicate that differences in response threshold are not a viable explanation for the PF observed by Aue et al. (2012). As mentioned in the introduction to this section, I anticipated that if a change in response threshold then the A-Br advantage observed by Aue et al. would be attenuated by the experimental manipulations and potentially reflected in the confidence ratings. Neither was the case. For Experiment 1, participants were not adopting a conservative response threshold for C-D pairs resulting in those responses being withheld. If that were the case, forcing participants to respond would have resulted in the output of the would-be responses. Nor were there any differences in the retrospective confidence ratings for the correct response, further indicating that there were no systematic differences in the quality of the responses being provided. Thus, the data indicate that C-D pairs are simply less well remembered than A-Br pairs and are not being attenuated by a conservative response threshold. In fact, Koriat and Goldsmith (1994, 1996) have demonstrated that adoption of a conservative threshold not only reduces the number of items recalled (as is the case for Experiment 2), but should be accompanied by more *accurate* responses when participants have the option to withhold a response. They define accuracy as the number of correct responses divided by the sum of correct and incorrect response. In other words, it represents the probability any output response is correct, as opposed to choosing to withhold a response.

With respect to Experiment 2, PF was observed despite asking participants to respond only with higher confidence responses. If participants were adopting a more liberal response threshold and, as a result, outputting more, lower quality responses, then asking participants to withhold low-quality responses should have selectively attenuated A-Br performance. This was not the case despite the fact that participants tended to be more confident in their retrospective confidence ratings relative to Experiment 1. Nor were there differences in the retrospective confidence ratings for A-Br and C-D pairs suggesting there was no systematic difference in the quality of response being provided for the two pairs.

In summary, there is no support a response threshold explanation in either the proportion of responses provided or in the retrospective quality of the response. Next I begin to investigate the second potential explanation for PF, the idea that A-Br pairs are better encoded relative to C-D pairs.

Tests of the Better Encoding Explanation

In the next section, I examine the better encoding explanation for PF posited earlier. To reiterate, I suggested that perhaps it is the case that during the study of A-Br pairs during list 2, participants are drawing on recent experience with individual items to store more complete version of the pair. Unfortunately, it is impossible to measure the contents of individual memory representations so the approach I adopted was to look for indirect effects of better encoding of the list 2 pairs. This was accomplished in different ways across experiments. For Experiments 3 and 4, I looked for indirect effects of better encoding by manipulating the components of the list 2 pair. In the experiments conducted thus far, and in Aue et al. (2012), both the cue and the target are repeated across lists for A-Br pairs (see Table 2 for an example). In Experiment 3 the cue and the target repetition across lists were manipulated independently. For A-D pairs, the cue (i.e., the face) was studied in both list 1 and list 2 but with targets (i.e., words) that were only on list 1 and list 2, respectively. Likewise, for C-B pairs a given target was studied with different cues on list 1 and list 2 that were exclusive to their respective lists. The rationale for the design is that if PF is a function of better encoding of A-Br pairs during the list 2 presentation, then the effect may be attenuated by pairing the repeated member of a pair with an unrepeated (i.e., weak) partner. The same approach is taken in Experiment 4, although list 1 consists of individual items rather than pairs of items. This technique was chosen because Aue et al. observed that manipulating the types of pairs in list 1 influenced the presence of PF. If it is the case that pairs are better encoded, then we would expect PF to be most robust when both the cue and the target are repeated across lists.

Experiment 5 sought to test the idea that if PF is the result of better encoding then the benefit should be unique to a test of list 2 memory. The rationale for Experiment 5 was that if the better encoding is occurring during list 2, when the individual items are presented a second time in a different pairing, and the new pair is stored as a separate memory trace from the original presentation, then the PF benefit should not extend to the memory for the pairs studied during list 1. To test the idea the same design was employed but instead of being tested on list 2, participants were tested for the memory of the list 1 pairs. If PF is observed for list 1 it would indicate that better encoding, as conceptualized thus far, would not account for PF.

The final test of the better encoding explanation was performed in Experiment 6 where a 15 min lag was added following study of list 2. The aim was to encourage participants to isolate the list 2 context when searching memory during test. The alternative is that participants could simply be relying on the test context to aid in memory search, which could reinstate parts of list 1 and list 2 depending on the match to the test context. Thus, if PF is still observed in a scenario when participants are encouraged to reinstate list 2 context, it would seem likely that it could be attributed to better encoding of the list 2 pairs.

Experiment 3: Manipulating list 2 pair strength

In Experiment 3, I began investigating the question of whether the repetition (i.e., strength) of items in A-Br pairs is driving the advantage observed by Aue et al. (2012). To accomplish this I manipulated whether each component of the pair, either the cue or target, was repeated across study lists.

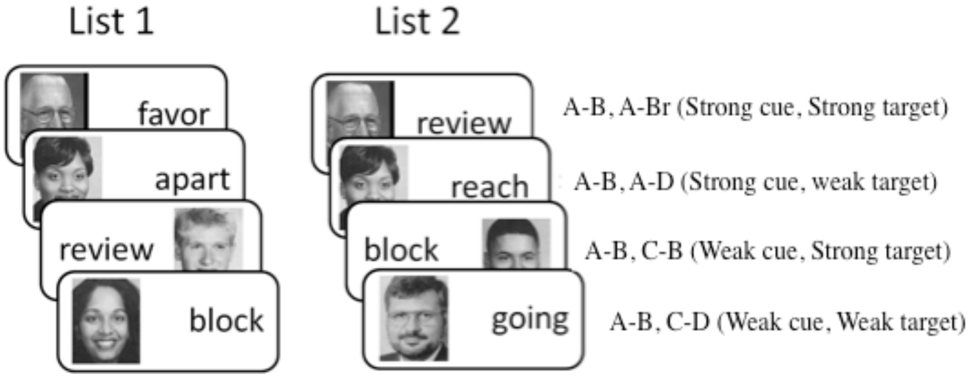

Figure 4. The general design of Experiment 3. Strong and weak refers to whether or not an item is repeated across study lists. The critical list pairs (list 2) were designed to orthogonally manipulate cue and target strength

As depicted in Figure 3, there were four types of pairs on the critical study list. In addition to A-Br and C-D pairs, I added pairs where the cue was studied on both lists and the target was exclusive to the second list (A-D) and pairs where the cue was unique to the critical list and the target was repeated across lists (C-B). Participant's experience of list 1 was the same as in the previous experiments. See Table 2 for a description of these pair types.

With respect to the hypothesized mechanisms for PF, the current design has the potential to answer important questions. As discussed earlier, it is not clear based on previous data what is being encoded better (e.g., items, associations), if anything at all. For instance, if A-Br pairs are better encoded during list 2 it could be because *both* items were studied in list 1. If it were the case then one would expect a robust interaction of cue and target strength reflecting an advantage for repeated items. However, if strengthening one member of the pairs (e.g., just the cue as in A-D pairs) is needed to prompt better encoding of the pair during list 2, then repeating individual items as either cues or targets should produce an advantage. Better encoding would lead to increases in correct responses since the encoding for the items and pairs should make them more distinct relative to other studied pairs and make targets more recoverable. Better encoding could also lead to fewer intrusions resulting from an increased likelihood of sampling the correct trace.

In regard to the response threshold and longer memory search explanations, if the advantage for A-Br pairs observed by Aue et al. (2012) is derived from the familiarity of the cue, then one would also expect the strength of the cue to influence performance. This would translate into more responses overall (both correct and incorrect) for pairs where the cue was repeated relative to pairs where it was not.

Method.

Participants. In total 73 participants from the Syracuse University subject pool took part in the experiment. Participants received course credit for their participation.

Materials. The materials were identical to those described previously.

Design. The details were the same as those described in Experiment 2. The major differences are the independent manipulation of cue and target repetitions described above and the fact there was no instructional manipulations nor were retrospective confidence ratings collected at test. Study time, test instructions, and encoding tasks are identical to previous experiments. For the list 2 pair type manipulation, instead of 12 pairs of each type during list 2 we had 6 pairs of each combination of cue and target repetition for a total of 24 items.

Results and Discussion. I separately examined correct and incorrect performance using a 2 (cue strength) x 2 (target strength) repeated measures ANOVA. I will consider the correct responses first which can be seen in the left panel of Figure 4. The interaction of cue and target repetition was not significant ($F(1, 72) = .022$, $p = .637$, $\eta_G^2 = .0007$). Despite the overall interaction being non-significant, planned comparisons of the A-Br pairs to C-D pairs did reveal that the two pair types differed significantly ($t(437) = 2.57$, $p = .01$, $d = .16$). This pattern replicates the basic results for correct performance of Aue et al. (2012). Next I examined the main effects present in the data.

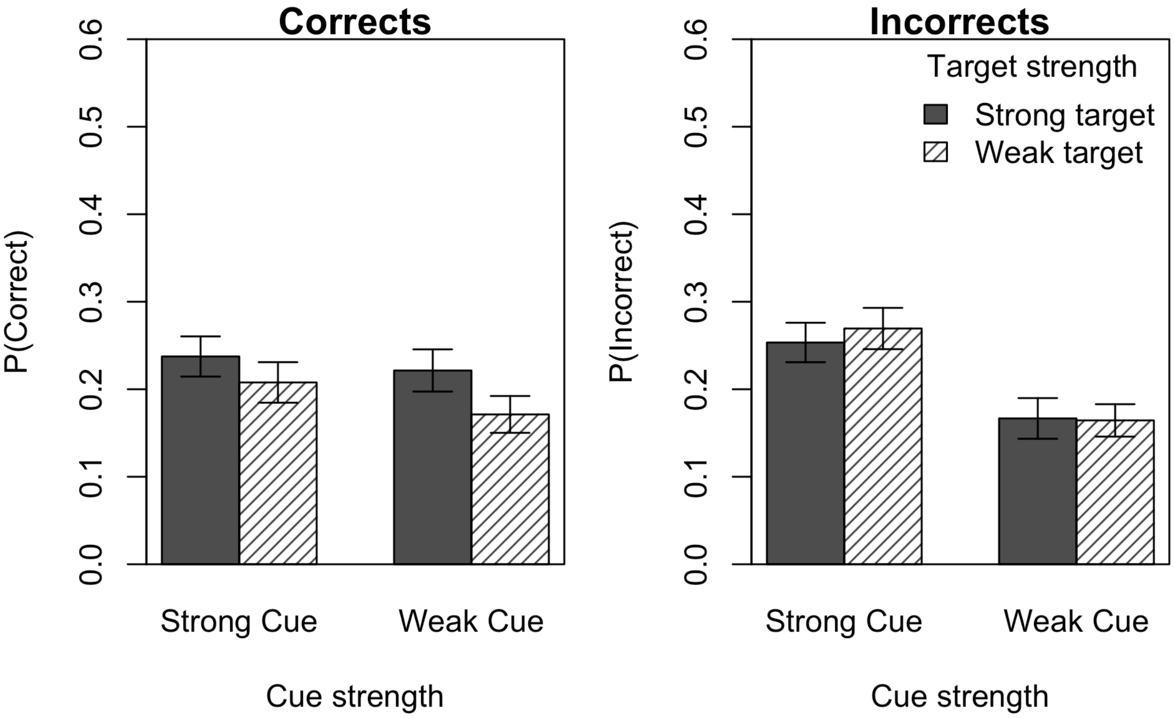

Figure 5. Correct (left panel) and Incorrect (right panel) responses for Experiment 3. A ‘strong’ cue/target was studied in both lists where as a ‘weak’ cue/target was studied only on list 2. Strong cues and strong targets, separately, tended to drive correct responses. Strong cues primarily drove incorrect responses.

I observed a trend towards an advantage for strong cues (*M* = .223, *SE* = .018) relative to weak cues (*M* = .196, *SE* = .017; *F*(1, 72) = 2.59, *p* = .112, η_G^2 = .0046). Participants tended to correctly recall the target more often when the cue had been studied in both lists regardless of the strength of the target. However, this is consistent with all three explanations for PF put forth. I also observed a main effect of target strength. Strong targets (*M* = .229, *SE* = .02) were recalled better relative to weak targets (*M* = .189, *SE* = .017; *F*(1, 72) = 4.19, *p* = .044, η_G^2 = .01). An advantage of target strength is only consistent with the better encoding hypothesis. During CR, the only information available at test is the cue and the participant would not have information about the target until a recall attempt was made. Thus, any modulation of response threshold or search time would likely be modulated by the cue.

Next I examined incorrect responses using an identical statistical design. Again the interaction of cue and target strength was not significant ($F(1,72) = .221$, $p = .639$, $\eta_G^2 = .0006$). However, a planned comparison of A-Br and C-D pairs demonstrated more intrusions for A-Br pairs relative to C-D pairs ($t(437) = 3.39$, $p < .001$, $d = .220$); a pattern that is evident in Figure 4 (right panel) and consistent with Aue et al. (2012). The main effect of target strength was not significant ($F(1,72) = .110$, $p = .740$, $\eta_G^2 = .0003$), however the main effect of cue strength was significant ($F(1,72) = 26.62$, $p < .001$, $\eta_G^2 = .062$). Participants intruded more often for pairs with strong cues ($M = .315$, $SE = .020$) relative to pairs with weak cue ($M = .222$, $SE = .020$; $F(1, 73) = 21.11$, $p < .001$, $\eta_G^2 = .062$). This is consistent with the hypotheses of a liberal response threshold or longer search explanations for A-Br pairs. If participants were engaging in either, the number of correct and incorrect responses would increase and would be driven by the familiarity of the test cue.

Next I examined the types of intrusions that participants made. The same chi-squared analysis was performed as for previous experiments, and the responses are summarized in Table 7. The A-Br pairs differed significantly from the expected values in terms of the distribution of intrusions across intrusion type (χ^2 (3, N = 88) = 490.76, $p < .001$). Qualitatively, intrusions for A-Br pairs tended to be the “List 1 partner.” The C-D pairs also differed significantly from the expected values (χ^2 (2, N = 54) = 19.99, $p < .001$). The A-D pair intrusions differed also significantly from the expected values (χ^2 (3, N = 96) = 153.37, $p < .001$) as did the C-B pairs (χ^2 (2, N = 61) = 50.75, $p < .001$).

The data from the current experiment do not clearly differentiate between the hypothesized mechanisms for PF. The strength of the cue drove both correct and incorrect performance in the data, consistent with both a response threshold account and a longer memory search account. Moreover, the cue driven nature of the incorrect responses is evidence against a better encoding account given that a strong cue should reduce intrusions by matching itself better and therefore leading to better sampling of the intended memory trace. Interestingly, however, correct performance was also driven by the strength of the to-be-recalled target. This would only be consistent with the better encoding for list 2 pairs hypothesis since target information would be unavailable when participants are a setting a response threshold or deciding how long to search memory. Next, I continued to investigate the impact of strengthening individual items, but change the composition of the first study list to single items instead of pairs.

Table 7. Experiment 3 intrusions by type

	A-Br		A-D		C-B		C-D	
	Expected	Observed*	Expected	Observed*	Expected	Observed*	Expected	Observed*
List 1 only	29.33	10	30.17	19	20.91	8	18.51	12
List 2 only	29.33	8	30.17	12	20.91	8	16.97	8
List 1 partner	2.44	36	2.74	22	NA			
List 1 and List 2	26.89	34	32.91	43	19.17	45	18.51	34

Note. $^{*}p < .05$; Expected values are based on the availability of an incorrect response from a given category for a given pair type.

Experiment 4: List 1 single items

In the following experiments, I manipulated the strength of the cue and target independently in a design that was free from proactive interference. I accomplished this by presenting individual items in the first study list where pairs had been presented in previous experiments. In Experiment 3 I observed that strong targets and, to a lesser degree, cues provided benefits at test. The question of import in the current experiments is whether the nature of previous experience (i.e., list 1 composition) with cues or targets matter for PF. Specifically, do the individual items need to be experienced in the context of a pair during list 1 in order to observe the benefit. It has been demonstrated that pairs of different types (e.g., word-word pairs versus word-face pairs) do not interfere with one another despite sharing similar features such as words and faces (Aue et al., 2012; Criss & Shiffrin, 2004, 2005). In this way, the current experiment will allow us to understand whether PF occurs from repetition of items rather than pairs and which component of the pair is contributing to the facilitation. As with Experiment 3, both the response threshold and longer search account would predict PF when the cue is more familiar (i.e., strong). The better encoding of list 2 account would predict a maximum advantage when both the cue and the target have been encountered previously, similar to A-Br pairs in previous experiments.

The experiment is similar in principle to Experiment 3 although it has been broken into a between subject and between list design, and a major source of interference (i.e., list 1 associations) removed in order to isolate the facilitation. The experiment was initially conducted as three separate experiments that, for ease of exposition, are discussed together. Across experiments participants always studied an initial list of single items. The single items were then used to create pairs where the cue was strong but the target was weak (e.g., A, A-D; Experiment

4a), the cue was weak but the target was strong (e.g., A, C-B; Experiment 4b), or *both* the cue and target were strong (e.g., A/B, A-B pairs; Experiment 4c). The pairs comprised half of the critical list and were accompanied by pairs unique to the second, critical list (e.g., C-D pairs). See Table 8 for a depiction of these pair types.

Method.

Participants. For the three experiments described, 41 participated in Experiment 4a, 42 in Experiment 4b, and 39 in Experiment 4c. All participants were recruited from the Syracuse University subject pool and received course credit for their participation.

Table 8. Types of pairs used in Experiment 4

Prior list study	**Potential critical list pairs**			
Study Item	Pair type	Cue	Target	Description
Absence	A/B, A-B	Absence	Hollow	Both cue and target are repeated
Pupil	A, A-D	Absence	Tissue	Repeated cue, new target
Hollow	A, C-B	Pillar	Hollow	New cue, repeated target
River	A, C-D	Pillar	Tissue	New cue, new target

Note. The above examples use only words for the studied pairs for expository purposes. In reality, the experiment employed both words and faces during the study lists.

Materials. The materials were identical to the previous experiments.

Design. During the initial study list, participants studied 24 individual items at a rate of 3 s each. Half of the items were words and the other half faces. As an encoding task, participants were asked to rate the pleasantness of the items on a nine-point scale. The second list was comprised of 24 pairs. Half of the pairs contained either one or two strong (i.e., repeated) items, manipulated across experiments, and the other half were pure weak (C-D) pairs. In Experiment 4a, the cue (face) was studied during list 1 as a single item but the target was exclusive to list 2. In Experiment 4b, the target (word) was studied in list 1 as a single item, but the cue was exclusive to list 2. In Experiment 4c, both the cue and target were studied in list 1 as single items. As with previous experiments, the type of pair was manipulated within subject. Which part of the list 2 pair (e.g., cue, target, or both) that was strong was the only aspect that differed across experiments.

Results and Discussion. To analyze the data I conducted three separate one-way repeated measures ANOVA for correct and incorrect responses for Experiments 4a, 4b, and 4c and the data are summarized in Figure 6. For experiment 4a where just the cue was experienced in list 1 as a single item, there was no impact of pair type on correct responses. Participants correctly recalled the target equally as often for A-D pairs (*M* = .254, *SE* = .024) as C-D pairs (*M* = .242, *SE* = .021; *F*(1, 40) = .169, p = .683, η_G^2 = .004). There was also no difference for the incorrect responses with A-D pairs (*M* = .199, *SE* = .026) and C-D pairs (*M* = .167, *SE* = .020; *F*(1, 40) = 1.56, p = .218, η_G^2 = .038) having a similar number of intrusions.

When just the target was experienced in list 1 as a single item, there were no differences between the performances on the pairs. Participants tended to correctly recall the target as often for C-B pairs (*M* = .254, *SE* = .026) as they did for C-D pairs (*M* = .222, *SE* = .025; *F*(1, 41) = 1.84, p = .183, η_G^2 = .043). There was also no difference between the C-B pairs (*M* = .148, *SE* = .028) and C-D pairs (*M* = .168, *SE* = .030; *F*(1, 41) = .907, p = .346, η_G^2 = .022) for intrusions.

For experiment 4c where both the cue and target were experienced in list 1 as single items, participants correctly recalled the target more often for A-B pairs (*M* = .340, *SE* = .032) relative to C-D pairs (*M* = .200, *SE* = .025; *F*(1, 38) = 31.61, $p < .001$, η_G^2 = .454). For incorrect responses, there was no difference between the A-B pairs (*M* = .162, *SE* = .024) and C-D pairs (*M* = .161, *SE* = .014; *F*(1, 38) = .002, *p* = .961, η_G^2 =.00006) for the number of intrusions.

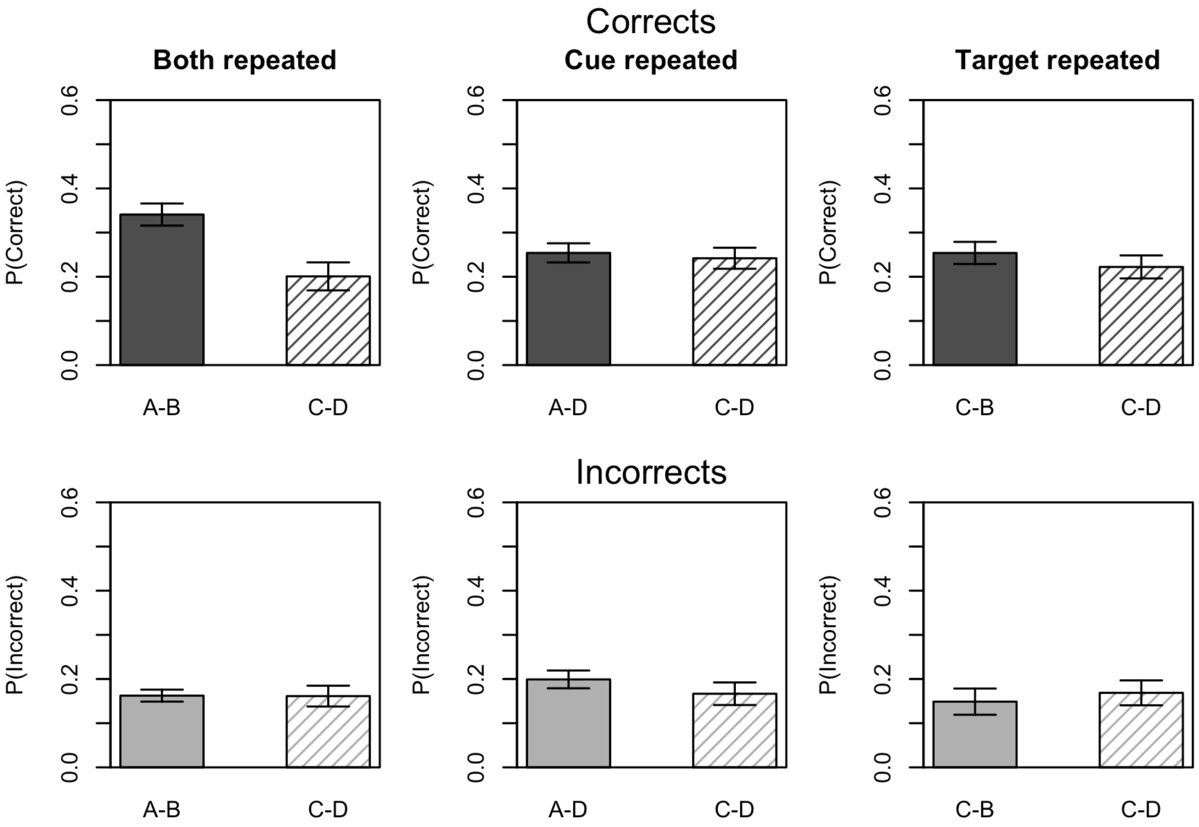

Figure 6. Performance across the three experiments summarized as part of Experiment 4. A large strength advantage was observed in correct responses when both the cue and the target appeared as individual items in list 1 (e.g., A, A-B; left panel) that was absent when just the cue appeared in list 1 (e.g., A, B, A-D; middle panel) or when just the target appeared in list 1 (e.g., A, B, C-B; right panel).

The intrusions for Experiment 4 are presented in Table 9. In general, there were fewer intrusions in Experiment 4 relative to previous experiments likely resulting from the reduced interference from list 1. We performed the same chi-squared analysis described before for each component of Experiment 4. All of the intrusion/pair-type combinations differed significantly from the expected values based on response availability. Table 9 provides the expected and observed values as well as the chi-squared statistic.

To summarize, the largest advantage for strengthening items was when both the cue and the target were strengthened by appearing as individual items in List 1. This pattern is consistent with the trend observed for Experiment 3 for correct performance to be highest for A-Br pairs. The fact that the advantage was isolated to when both the cue and target were repeated could be consistent with the better encoding explanation for PF if we assume that the individual items were better encoded when they were associated during list 2. Further, the results are inconsistent, or at least constrain, response threshold or longer memory search accounts given that strong cues alone did not increase correct or incorrect responding. It could be the case that participants rely on the familiarity of the cue having been experienced as part of a pair to change their response threshold or initiate a longer memory search. However, such an explanation alone would not be able to account for the observed advantage of strong cues and targets. Likewise, one might have also expected strong targets to have better performance as they did in Experiment 3; however, that was not the case.

Table 9. Intrusions by pair type for Experiment 4

	A-B		C-D	
	Expected	Observed*	Expected	Observed*
List 1 only	NA		NA	
List 2 only	30.26	16	27.26	19
List 1 and List 2	27.74	42	29.74	38
χ^2		14.05		4.80

	A-D		C-D	
	Expected	Observed*	Expected	Observed*
List 1 only	18.51	3	16.46	6
List 2 only	35.48	51	31.54	42
List 1 and List 2	NA		NA	
χ^2		19.78		10.11

	C-B		C-D	
	Expected	Observed*	Expected	Observed*
List 1 only	NA		NA	
List 2 only	22.96	10	24.87	17
List 1 and List 2	21.04	34	27.13	35
χ^2		15.29		4.77

Note. * $p < .05$

Experiment 5: Test of list 1 associations

The next experiment examined the retroactive effects of repeating items across lists in different pairs. Everything that has been discussed up until this point has examined proactive interference and facilitation where the aim is to examine the influence of previously studied pairs (i.e., A-B pairs) and items on the recall of more recent pairs (e.g., A-Br). In a retroactive design, the critical list shifts from the second list to the first list. This design allowed me to examine the influence of subsequent learning on previously learned information. Thus, the study scenario is identical to Aue et al. (2012; Experiment 1) and of Experiments 1 and 2 in the current paper. However, at test participants are provided with a cue and asked to recall the target with which the item was first studied. The primary question of interest here was whether or not PF, or in this case retroactive facilitation (RF), was present for first list associations[5]. With respect to the proposed hypotheses, both response bias and longer searches would predict that the A-Br advantage should persist given that the cue provided at test is more familiar and would induce either longer searches or a shift in the response threshold. However, if the pairs are being encoded more completely during their *second* presentation and they are stored in a separate trace from the list 1 pairs, asking participants to recall the associate from the initial study list should eliminate the advantage for correct responses.

Method.

Participants. A total of 57 Syracuse University subject pool members participated in the experiment. Participants received course credit for their participation.

Materials. The materials were identical to previous experiments.

---

[5] RF has also been observed elsewhere in slightly different designs (e.g., Bruce & Weaver, 1973; Robbins & Bray, 1974; Tulving & Watkins, 1974).

Design. The details were the same as those described in Experiment 2. The only changes are the fact there was no instructional manipulation was performed at test, nor were confidence ratings collected. As described above, participants were tested for their memory for the first list. At test, participants were shown a face and asked to recall the word that it was studied with during the *very first* list. The pairs are created in a manner identical to previous experiments. As can be seen in Figure 1, in Aue et al. (2012) half of the pairs on list one were comprised of items that were subsequently rearranged into new pairs on list 2 (i.e., A-Br, A-B). Also on list 1 are pairs that are exclusive to list 1 (i.e., C-D, A-B). As such we will use the same terms to refer to these pairs as we have in the discussion of other experiments.

Results and Discussion. A one-way repeated measures ANOVA was performed separately for correct and incorrect responses. I examined correct responses first. Despite overall performance being worse relative to previous experiments, participants tended to respond correctly more often to A-Br pairs (M = .146, SE = .017) relative to C-D pairs (M = .105, SE = .014; $F(1, 56) = 5.65$, $p = .021$, $\eta_G^2 = .092$). Further participants also responded incorrectly more often to A-Br pairs (M = .377, SE = .029) relative to C-D pairs (M = .209, SE = .024; $F(1, 56) = 42.32$, $p < .001$, $\eta_G^2 = .430$), consistent with Experiments 1-3 in the current paper and those of Aue et al. (2012).

Incorrect responses were coded as they have been in previous experiments with one exception. The list 2 partner of the cue “A” took the place of the list 1 partner intrusion code used in previous experiments. To examine whether the pattern of intrusions differed from the available within-experiment responses, we performed the same chi-squared analysis as before. Both A-Br (χ^2 (3, N = 190) = 1658.28, $p < .001$) and C-D pair (χ^2 (2, N = 102) = 36.21, $p < .001$)

intrusions differed significantly from the expected values generated based on available responses. The frequencies of intrusions by type are provided in Table 10.

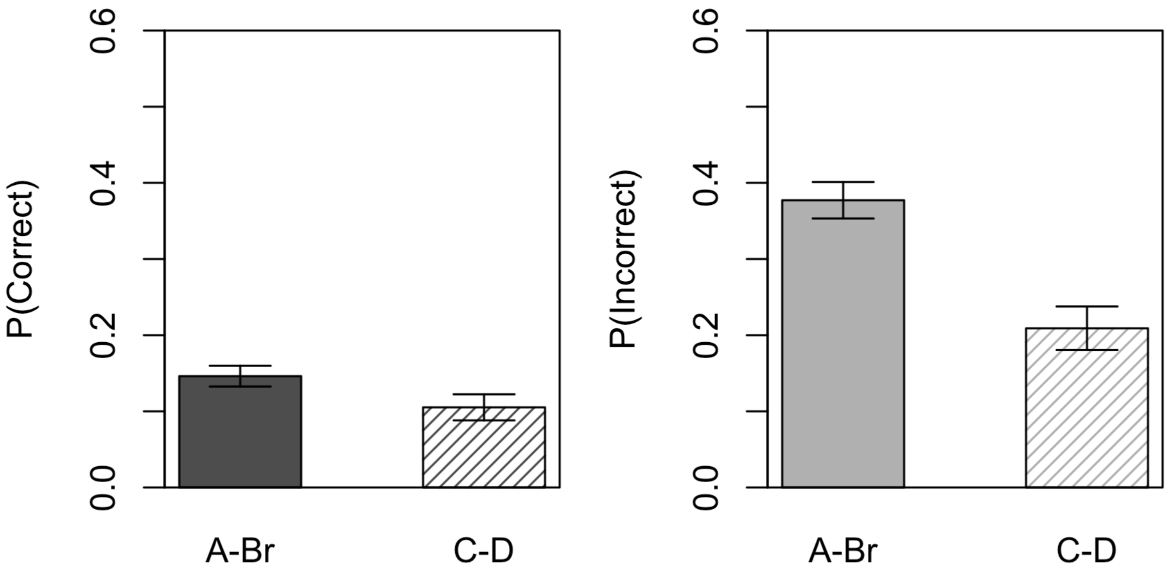

Figure 7. When participants were tested on their memory for the initial study list in a retroactive interference design, retroactive facilitation was observed relative to pairs exclusive to the initial study list.

The fact that RF was observed for A-Br pairs indicates that storing a more complete representation for items repeated across lists cannot alone explain the data if participants are storing the items in a separate trace from the list 1 presentation. We will discuss this idea in greater depth in the discussion. If it were the case, participants should not demonstrate a RF advantage for list 1 associations given that the advantage would be conferred during list 2 presentations. However, the response threshold and longer memory search accounts may be able to account for the data given that the cue is still more familiar from having been presented twice.

Table 10. Intrusions by type for Experiment 5

	A-Br		C-D	
	Expected	Observed*	Expected	Observed*
List 1 only	63.33	11	32.06	9
List 2 only	63.33	27	34.97	32
List 2 partner	5.28	97	NA	
List 1 and List 2	58.06	55	34.97	61

Note. *$p < .05$

Experiment 6: Expanding the retention interval

As discussed, the results of Experiment 5 are potentially difficult to reconcile with the better encoding hypothesis. This is because if it were simply that the List 2 pairs were better encoded then one would not expect the benefit to extend retroactively to List 1. However, if participants are spending more time searching for the target in response to a familiar cue a benefit for either test could be observed. The results, thus far, have demonstrated that repeating the cue alone does not differentially influence proactive facilitation. In the current experiment, I attempted to put the familiarity of A-Br cues and C-D cues on closer to equal footing by adding a 15 min retention interval between list 2 and final test. To accomplish this, I employed the same general design as previous experiments with the addition of a 15 min delay between list 2 study and test. The goal of adding the retention interval was to encourage participants to reinstate the list 2 context at test. Dennis and Humphreys (2001) suggest that for shorter study sessions (in their case it was specifically short list lengths) at test participants simply use the current test context to reinstate the study scenario. In the current design, this would likely result in substantial overlap between list 1 and list 2 reinstatement resulting in greater familiarity for the repeated cues in A-Br pairs relative to the weak cues in C-D pairs. Adding a delay after study encourages participants to attempt reinstating the necessary study context (i.e., list 2 context). Of the pairs associated with the isolated context, both A-Br and C-D pairs appear once. In this scenario if cue familiarity were the only feature driving PF then we would expect the difference between A-Br and C-D pairs to be attenuated. If, however, PF is the result of better encoding of the pairs during list 2 then we would expect the advantage to persist.

Method. There were two components to the experiment. The first was a replication of Aue et al. (2012) with shorter lists. The second was the same design with the addition of a 15 min delay following the study of List 2. The experiment will be discussed together, and differences will be noted.

Participants. A total of 53 participants took part in the "No Delay" experiment, while 56 took part in the 15 min delay experiment. Recruitment and compensation was the same as previous experiments.

Materials. The materials were identical to previous experiments with the exception of a puzzle that participants in the 15 min delay condition completed during the retention interval.

Design. The design of the experiment was nearly identical to that of Aue et al. (2012). The within subjects manipulation was the pair types studied during List 2 (e.g., A-Br and C-D; Table 2). The primary difference was a between subject manipulation of retention interval. In the 0 min retention interval, after studying list 2 participants completed a 60 sec arithmetic task and then went immediately into the cued recall task as in Aue et al. In the 15 min delay condition participants completed the same 60 min arithmetic task. Afterwards, participants were prompted that there would be a brief delay before proceeding and to please see the experimenter. Participants were then seated at a table with anyone else currently taking the experiment and asked to work together on the puzzle. After 15 min had passed they were taken back to their booth to complete the cued recall task as before.

In addition to the delay, I also anticipated that performance would be worse following the delay. Given that performance for the task was already poor I decided to shorten the list to 16 items from the 24 Aue et al. (2012) used in an attempt to prevent a floor effect. Everything else about the experiment is identical to previous experiments.

Results and Discussion. I employed two-way mixed ANOVA with delay (between subjects; 0 min vs. 15 min) and pair condition (within subject; A-Br vs. C-D) for the correct and incorrect responses separately. For correct responses, I observed main effects of both delay and pair condition. For delay, as one might expect performance was better for the short delay condition (M = .314, SE = .027) relative to the long delay condition (M = .164, SE = .0193; $F(1, 107) = 20.76$, $p < .001$, $\eta_G^2 = .128$). For the pair condition, replicating the results of Aue et al. (2012) A-Br pairs (M = .266, SE = .028) were better recalled than C-D pairs (M = .208, SE = .019; $F(1, 107) = 10.01$, $p = .002$, $\eta_G^2 = .022$). For incorrect responses, I observed only a main effect of pair condition. Across the delay, participants made roughly the same amount of incorrect responses for the short delay condition (M = .258, SE = .020) as they did for the long delay condition (M = .267, SE = .024). For the pair condition, consistent with Aue et al. I observed more intrusions for A-Br pairs (M = .297, SE = .019) relative to C-D pairs (M = .228, SE = .019; $F(1, 107) = 9.35$, $p = .003$ $\eta_G^2 = .028$). The data are plotted as a function of delay and pair type in Figure 8.

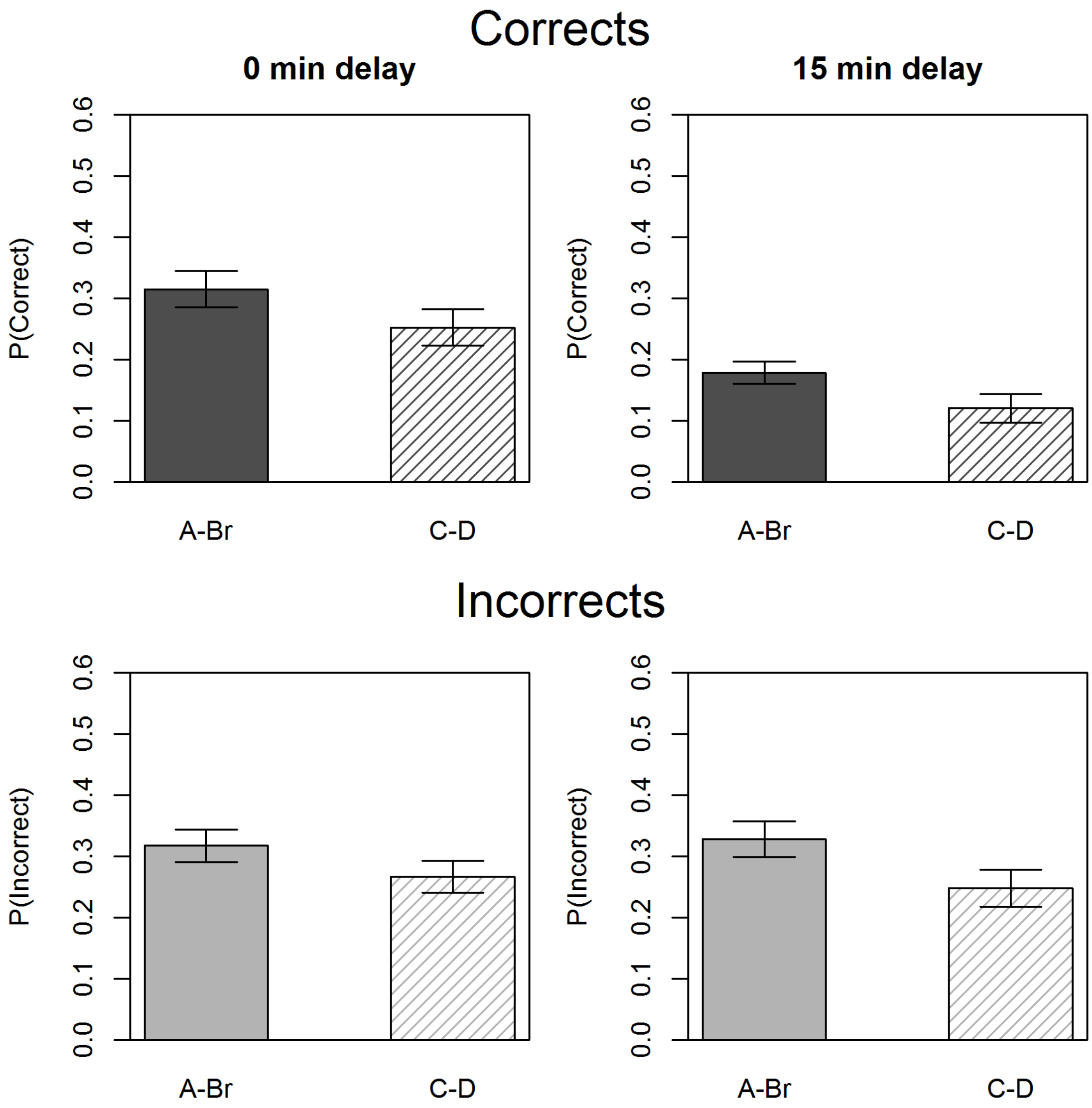

Figure 8. Memory performance in Experiment 6 following a 0 min delay (left panels) or a 15 min delay (right panels). The A-Br advantage observed by Aue et al. (2012) persist across retention intervals.

Next I examined the types of intrusions that participants tended to make. The data are presented in Table 11. The proportions of intrusions across types for A-Br pairs in the 0 min delay group differed significantly from the expected values based on response availability (χ^2 (3, N = 93) = 463.88, *p* < .001). The intrusions for C-D also differ significantly from the expected values (χ^2 (2, N = 72) = 46.35, $p < .001$). The same was true for the 15 min delay group. The distributions of both the A-Br pairs (χ^2 (3, N = 87) = 480.83, $p < .001$) and the C-D pairs (χ^2 (2, N = 49) = 25.87, $p < .001$) differed significantly from the expected values.

In summary, the persistence of the A-Br advantage provides further support to the better encoding explanation of PF. The purpose of the delay was to make the cue less familiar. The cue is critical for cued recall because it is all that the participant has other than their internal context with which to search memory. The rationale for the current experiment was that if the List 2 pair was better encoded during the second study list then any advantage would persist across the long retention interval.

Discussion of the Better Encoding Results

Across Experiments 3-6, the idea that list 2 pairs were being better encoded as a result of the individual items being presented in list 1 in different pairings. This idea was tested in multiple ways. First, in Experiments 3 and 4, by manipulating the composition of the pairs when they are studied during list 2. Then in Experiment 5 by testing participants for their memory of pairs on list 1, which by design could not have been affected by better encoding on list 2. Lastly, in Experiment 6, better encoding was tested by encouraging participants to reinstate the list 2 context at test by adding a long delay between study and test.

Experiments 3 and 4 both demonstrated that correct performance (i.e., PF) was highest when both the cue and the target were studied in list 1 (i.e., A-Br pairs), relative to when they were not (i.e., C-D pairs). Likewise, for Experiment 6 I argue the persistence of the A-Br advantage across the 15 min delay is evidence of better encoding given. However, there are two points that are difficult to reconcile with a better encoding account. First, the data from Experiment 5 contradict the narrative that better encoding is occurring during list 2 study given that the PF benefit extended retroactively to memory for list 1 associations. Second, as mentioned before, a better encoding explanation alone would not also account for the PI observed in nearly every experiment. That being said these data could still be explained by a longer search explanation, which I examine next. However, if not better, then differential encoding may be able to explain these data. Aue et al. (2012), also briefly suggested that transient associations between studied items might be formed while encoding as a mechanism for PF. One could imagine transient associations taking the form of associating the list 2 response for a given word-face pairing to the initial word-face pairing. Thus after studying A-B in list 1, when you see A-D in list 2 you might store A-B-D. This idea is similar to one put forth by Wahlheim & Jacoby (2013) where the details have been fleshed out somewhat more and paired with the idea that this only occurs when you remember seeing the individual items in list 1, a phenomenon called study phase retrieval. If so, this would explain the retroactive facilitation observed in Experiment 5 and perhaps the PI, especially considering that many of the intrusions come from the list 1 partner. I return to the question in the General Discussion and provide a fuller analysis of the idea.

In the final experiment, Experiment 7, I address the question of whether participants are searching memory longer for A-Br pairs by examining response times during recall of A-Br and C-D pairs.

Table 11. Intrusions by type for Experiment 6

	0 min delay			
	A-Br		C-D	
	Expected	Observed	Expected	Observed*
List 1 only	31.00	5	25.04	14
List 2 only	31.00	18	21.91	6
List 1 partner	3.88	45	NA	
List 1 and List 2	27.13	25	25.04	52
	15 min delay			
	A-Br		C-D	
	Expected	Observed*	Expected	Observed*
List 1 only	29.00	5	17.04	7
List 2 only	29.00	11	14.91	8
List 1 partner	3.63	44	NA	
List 1 and List 2	25.38	27	17.04	34

Note. * $p < .05$

Testing Longer Memory searches

The primary aim of Experiment 7 was to test whether participants are spending more time searching memory for A-Br relative to C-D pairs. As discussed before, this would be akin to encountering someone who seems familiar, but whose name you are unable to recall. As a result, you might spend more time trying to remember the person's name relative to when the person seems less familiar. I hypothesized that the same behavior could potentially explain PF. If the cues for the A-Br pairs seemed more familiar to participants, this could encourage them to search memory longer for the pair relative to C-D pairs. It is reasonable to assume that they would be more familiar given that the individual items in A-Br pairs tend to be better recognized relative to C-D pairs (Aue et al., 2012; Criss & Shiffrin, 2005). In memory models such as REM (Diller, Nobel, & Shiffrin, 2001) the time spent searching memory is represented as an index of the number of times the model attempts to retrieve information from memory and either fails or rejects the retrieved information. Thus, each additional attempt at searching memory affords an additional opportunity to retrieve either the correct memory trace, retrieve an incorrect memory trace, or to fail to retrieve the memory. As such, having an additional opportunity to retrieve something from memory could boost the likelihood of retrieving either the correct or incorrect trace for A-Br pairs relative to C-D pairs and could potentially explain both the PF and PI observed in the current data.

Experiment 7: Proactive effects on reaction time

If it is the case that participants are spending more time searching memory for A-Br pairs relative to the equivalent of C-D pairs, then that could manifest as an increased latency for participant’s initial signal of readiness to respond. In other words, participants should take longer to respond to A-Br pairs relative to C-D pairs. In particular, I expected participants to provide a greater number of slower responses for A-Br pairs relative to C-D pairs. This effect could manifest for correct or incorrect responses, as well as for the duration at which participants ultimately terminate their memory search. I also anticipate that participants would have longer RTs for incorrect relative to correct responses as has been observed elsewhere (e.g., Nobel &

Shiffrin, 2001). This pattern would reflect both the PF and PI aspects of the Aue et al. (2012) data.

Alternatively, if the PF aspect of Aue et al. (2012) is driven by better encoding of list 2 pairs, then one might expect to observe strength effects in RT wherein correct responses to A-Br (i.e., strong) pairs are faster than for C-D (i.e., weak pairs). Such differences in RT have been observed in traditional strength paradigms (i.e., a pair is repeated multiple times) in CR (e.g., Guez & Naveh-Benjamin, 2006) and associative recognition (e.g., Verde & Rotello, 2004). However, as noted these studied have involved the repetition of pairs rather than items so how the RTs will bear out if list 2 is better encoded is speculative.

Method.

Participants. A total of 40 participants took part in the experiment. Participants received course credit for their participation.

Materials. The stimuli used in the experiment were same as those employed in previous experiments. Unique to this experiment was the use of MATLAB (R2011a, MathWorks) for stimulus presentation and response timing.

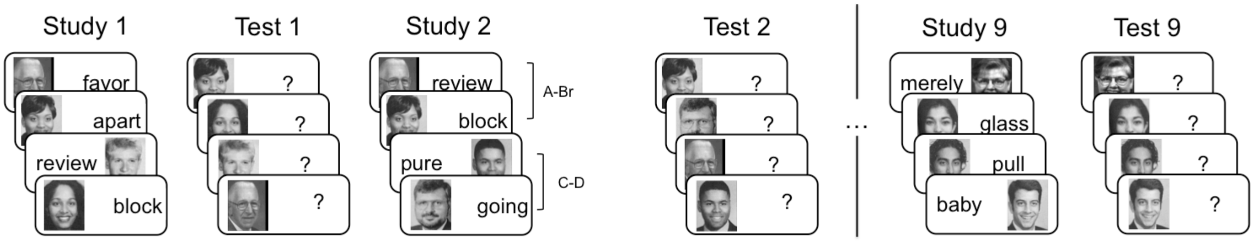

Figure 9. The design for Experiment 7. Participants went through 9 study-test cycles. For the second through ninth cycle participants studied half A-Br and half C-D pairs where A-Br were comprised of rearranged C-D pairs from the immediately preceding list

Design. Given that reaction times are notoriously noisy I adjusted the design of the experiment in an attempt to increase the number of observations. To improve performance I used the shorter list length described in Experiment 6. I also added multiple study-test cycles to increase the number of responses per participant overall. The general design is presented in Figure 9. During the experiment, participants completed nine study-test cycles with lists of 16 word-face pairs and 16 test trials. With the exception of the first study-test cycle, each list contained half A-Br pairs and half C-D pairs. Pairs were studied for 3 s with a word and face appearing side by side. The cue-target presentation location (i.e., left, right) was randomized. Participants were asked to engage in one of the two previously described encoding tasks for each list. To reduce prior list intrusions, the encoding tasks alternated for each studied list such that a participant never completed the same encoding task for two consecutive lists.

The first study cycle contained all C-D pairs comprised of words and faces. Participants were then tested on their memory for the first list using CR. On each test trial, participants were shown a face and asked think of the word that it appeared with on the most-recent list. Once they had thought of the target word or decided that they do not know it they were asked to indicate as much by pressing the space bar. This measure is the primary quantification of reaction time that I will discuss referred to as *Signal RT*. Next the cue disappeared and participants were prompted to type out their response or to type 'no' if they did not recall the target. At this point I collected two additional reaction time measures. Acknowledging that participants could still be searching memory while typing in a response, I first measured when participants ultimately made the *first keystroke (First KS)* of their response calculated as the duration between their Signal RT and the timestamp for the first key press of the response. I also measured the *last keystroke (Last KS)* of the participant's response, when they ultimately press 'enter' to submit their response. From these measures I also examined the time spent entering a response (i.e., *response duration*). After each test list participants had a 60 s break prior to being prompted to proceed to the next study-test cycle. During the break, participant completed a 60 sec arithmetic task that has been described previously. There were a total of nine study-test cycles lasting approximately one hour.

To analyze the reaction time data, I used the R package "retimes" (Massidda, 2013) for fitting the reaction time data to an ex-Gaussian distribution. Memory performance was analyzed using the techniques described previously.

Results and Discussion.

Memory performance. For the nine study-test blocks that participants completed, only the last eight are considered in the forthcoming analyses given that the first list contained only C-D pairs. Additionally, I collapsed across test blocks in the forthcoming analyses because there was no interaction with, nor main effect of test block for correct or incorrect performance. To analyze the data, I employed a one-way repeated measure ANOVA for performance between the pair condition group. For correct responses I observed a significant A-Br advantage in cued recall. Participants tended to correctly recall A-Br pairs (M = .462, SE = .037) more often than C-D pairs (M = .411, SE = .036; $F(1, 39) = 17.83$, $p < .001$, $\eta_G^2 = .312$). Interestingly, for incorrect responses[6] the proactive interference results that we had observed previously were conspicuously absent. Indeed participants responded incorrectly to A-Br pairs (M = .163, SE = .022) as often as they did to C-D pairs (M = .163, SE = .021; $F(1, 39) = .008$, $p = .964$, η_G^2 =5e-5). These data are an interesting result that has precedence. Szpunar, McDermott, and Roediger (2008) noted that

---

[6] I chose to omit the analyses on intrusions provided for previous experiment chiefly because the number of available responses used to generate the expected values for previous experiments grows with each subsequent list in the current design. Additionally the intrusion categories do not map onto the previous categories, nor were there differences in the overall intrusion rate for A-Br pairs relative to C-D pairs so such an analysis may not be particularly informative.

testing provides a release from the build up of proactive interference with participants who did not receive intervening tests providing an increasing number of intrusions as study-test cycles increase. Some have theorized the release to be driven by a test-driven context shift (Jang & Huber, 2008; Klein, Shiffrin, & Criss, 2007; Lehman & Malmberg, 2012). The current data indicate that PF may rely on a separate mechanism from PI.

Response time. Next I examined response times for the correct, incorrect, and no recall responses. As stated earlier, if participants are spending more time searching memory for A-Br pairs relative to C-D pairs then this may be reflected in the time that it takes participants to output a response. To analyze the data I used a distributional analysis (e.g., Heathcote, Popiel, Mewhort, 1991; Rouder & Speckman, 2004) fitting each reaction time distribution to an ex-Gaussian distribution using maximum likelihood estimation[7]. The ex-Gaussian distribution is a convolution of a Gaussian distribution and an exponential distribution and tends to provide excellent fit to response time data (Hohel, 1965; Heathcote, Popiel, Mewhort, 1991; Matzke & Wagenmakers, 2009). There are three parameters that influence the shape of the distribution: the mean (μ) and standard deviation (σ) of the Gaussian component which shift the distribution along the x-axis and change the spread of the distribution, respectively, and the mean of the exponential component (τ) which influences the thickness of the tail of longer distributions. If it is the case that participants are spending more time searching for A-Br pairs relative to C-D pairs we might expect A-Br correct responses to have a higher μ relative to C-D pairs. It could also be the case that the change in the distribution is found in the tail for A-Br pairs indicating a greater

---

[7] Reaction time data was truncated by eliminating responses lying beyond ±2 standard deviations from the mean RT for a given condition. The SIMPLEX method was used during MLE. Confidence intervals on the parameter values were generated through 10,000 bootstrapped samples.

abundance of longer responses. Such a pattern would be represented as differences in τ. I collected data for the three response time measures: Signal RT, First keystroke RT, and Response duration. Only the data for Signal RT is presented here as the data for First keystroke RT and Response duration replicated the qualitative findings. Analyses and a brief discussion of the First keystroke and Response duration analyses are provided in Appendix A.

Participants were instructed that once a face appeared on screen they were to attempt to recall the response in its entirety then signal through key pressing the spacebar that they were prepared to enter a response. The latency between stimulus onset and key press is Signal RT. The response distributions for the three response types are depicted in Figure 10 and the descriptive statistics are provided in Table 12. As is evident from the figure for all response types, there is substantial overlap in the distributions. Focusing on correct responses in particular, the most notable difference between the distributions is the fact that the C-D pairs have a slightly longer and thicker tail relative to A-Br pairs. This pattern indicates that there tended to be more correct responses at longer response times for C-D pairs relative to A-Br. The same was true for incorrect responses although to a lesser extent. In addition, for incorrect responses, A-Br pairs have a slightly greater density at shorter RTs than C-D pairs as evidenced by the higher peak in the middle panel of Figure 10. The qualitative differences in the intrusion distribution are interesting given the absence of proactive interference in the memory performance data. For no recall responses, A-Br pairs tended to have a more drawn out tail, again indicating the presence of longer responses. In other words, participants were occasionally taking longer to decide to terminate their memory searches for the A-Br pairs.

The model was fit to aggregated and the results of the model fitting are presented in Figure 11. In general, the ex-Gaussian distribution captured the pattern of data well for both A-

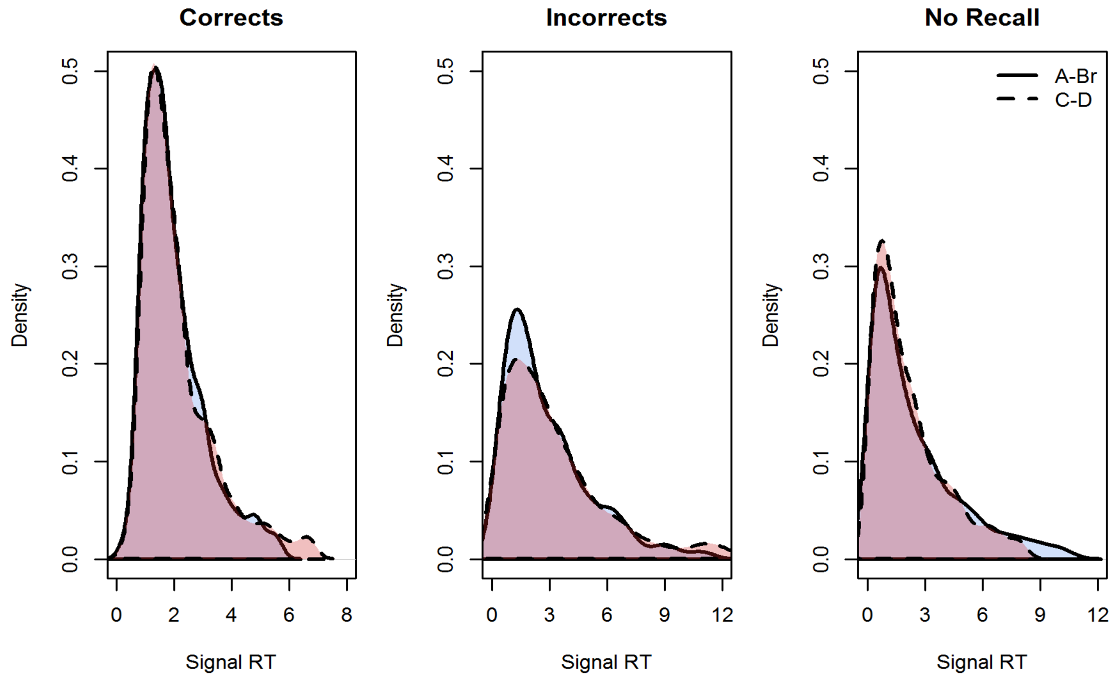

Figure 10. Presented are the response distributions for Signal RT of A-Br (solid line; blue distribution) and C-D (dashed line; red distribution) responses for each of the three response types. The purple area on the plot denotes the overlap between the two distributions.

Br and C-D pairs. The descriptive statistics for the behavioral data and the best-fitting ex-Gaussian parameters estimates and 95% confidence intervals of bootstrapped parameters estimates are provided in Table 12. As can be seen in Table 12, for the behavioral data the mean RT for correct responses are similar for A-Br and C-D pairs. However, when we examine the results of the bootstrapped model fitting, we can see that the largest difference in parameter values is for τ where C-D pairs tended to have a larger value for both correct and incorrect responses. Moreover, if we use the overlap of the 95% CI of the ex-gaussian parameter estimates as a rough indicator of credible differences between the distributions C-D tended to have a larger τ parameter estimate when a response was generated; although the difference is marginal for corrects. Given that τ is the parameter for the exponential component of the distribution and influences the shape of the tail, it would seem that participants are spending *less* time to retrieve incorrect, and to a lesser extent correct, responses for A-Br pairs relative to C-D, contrary to my expectations. However, for No recall responses, the opposite was true. The A-Br pairs tended to have a larger value of τ relative to C-D pairs indicating that participants hesitated longer before entering their ‘no’ response for A-Br pairs, as predicted. For incorrect responses, τ was larger, and the CIs different, for C-D pairs relative to A-Br pairs while μ was larger for A-Br pairs relative to C-D pairs. This pattern reflected the greater density at faster RTs for A-Br pairs discussed earlier, which drew the tail of the distribution in relative to C-D pairs.

Table 12. The best-fitting ex-Gaussian parameters for Signal RT

	Behavioral data Descriptives		Ex-Gaussian Model fits to behavioral data		
	M	*SE*	μ	σ	τ
Correct					
A-Br	1.89	.084	.894 (.839, .946)	.300 (.237, .341)	1.09 (1.02, 1.17)
C-D	1.88	.075	.880 (.838, .927)	.252 (.204, .293)	1.25 (1.16, 1.34)
Incorrect					
A-Br	3.25	.285	.300 (.256, .572)	.080 ($1.12e^{-7}$, .333)	2.57 (2.22, 2.76)
C-D	3.38	.259	.278 (.210, .362)	.069 ($1.18e^{-7}$, .134)	3.26 (2.93, 3.61)
No Recall					
A-Br	2.31	.156	.128 (.095, .145)	.030 ($1.35e^{-6}$, .041)	2.35 (2.20, 2.50)
C-D	2.26	.161	.135 (.122, .164)	.041 ($2.92e^{-7}$, .053)	2.05 (1.94, 2.16)

Note. 95% confidence intervals generated based on 10,000 bootstrapped samples.

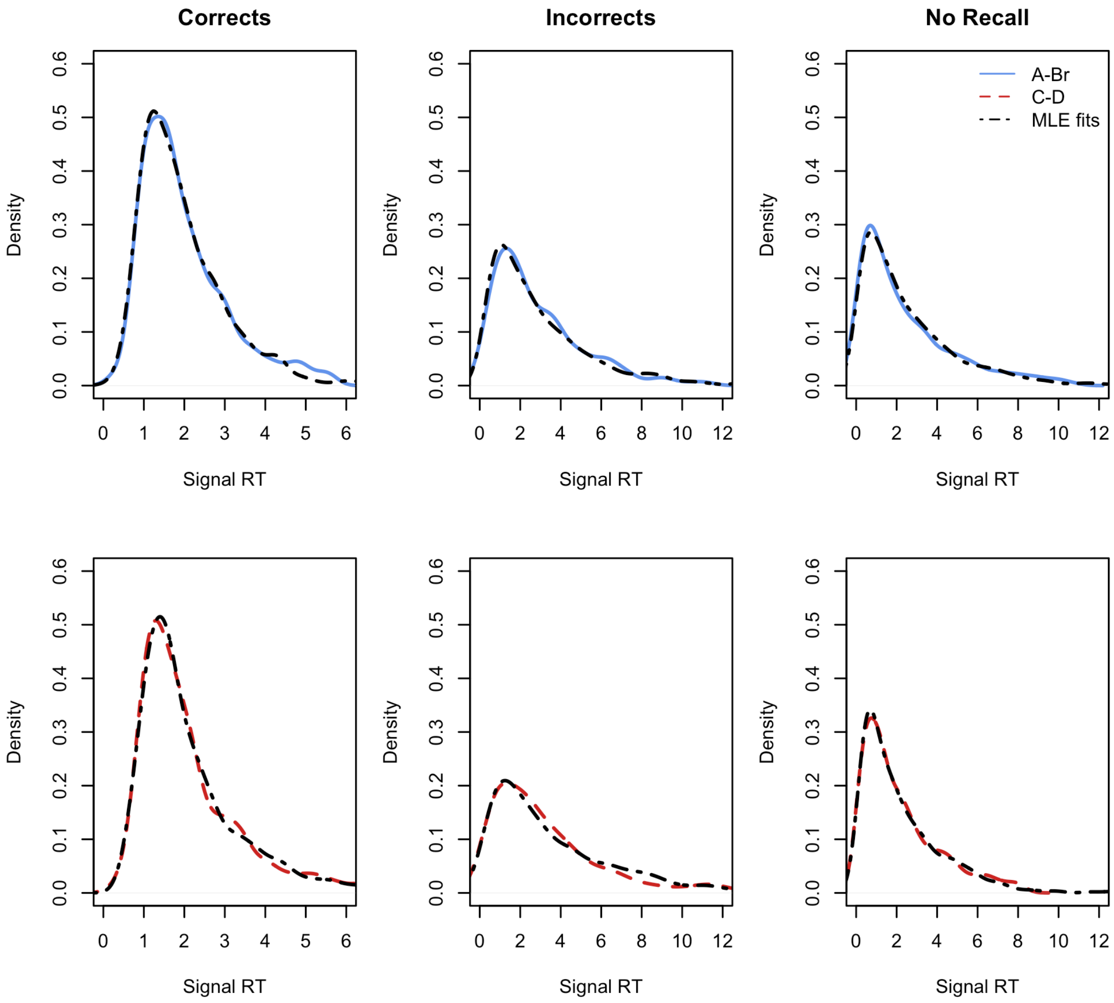

Figure 11. The top row contains the data for the A-Br behavioral data and model fits while the bottom row contains C-D pair behavioral data and model fits. Fit statistics and parameter values for each distribution are provided in Table 12.

Lastly, I examined how performance changes for the pair conditions as a function of response time. This is an approach typically employed when examining predictions made by sequential sampling models of reaction time, but I am simply using it an additional data visualization tool here. The data are presented in Figure 11. In the figure, the degree to which the quantiles for the A-Br and C-D pairs are shifted relative to one another along the y-axis indicates differences in RTs whereas the differences along the x-axis indicate differences between the groups on memory performance. For correct responses, the largest difference in memory performance between A-Br and C-D pairs is at the fastest quantiles where the A-Br pairs tend to sit to the right of the C-D pairs. Notably, however, there are not RT differences between these quantiles as they occupy roughly the same space along the y-axis. We also see differences between A-Br and C-D pairs in the No Recall data wherein No Recall responses tend to be less common for A-Br pairs at the faster quantiles. This may reflect a hesitancy to reject the A-Br pairs associated with a longer memory search.

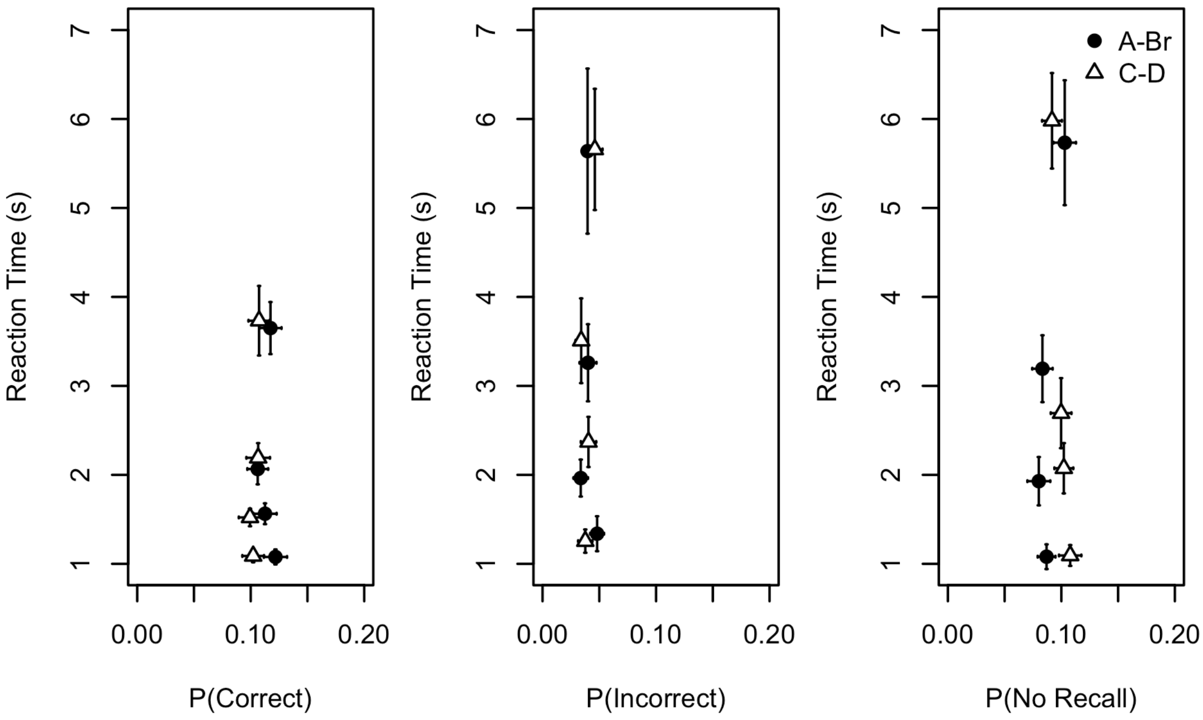

Figure 12. Median reaction time for quantiles plotted as a function of the proportion of responses for the given quantile.

To summarize, I did not observe strong evidence of longer searches leading to more correct responses for A-Br relative to C-D pairs. Through the ex-Gaussian model fitting, I did find that the best-fitting parameters for no recall responses fit a distribution that contained some longer responses for A-Br pairs. This pattern is consistent with the idea that participants were taking longer to terminate their memory search. However, these data alone do not explain the advantage that we observe for correct responses. For correct responses, A-Br pairs had fewer responses at longer intervals that, if anything, are more consistent with a strength interpretation of those responses. This was observed in spite of the interfering nature of the experimental design. Research examining strength effects (although in a slightly different design) in cued recall have found that repetition improves performance and decreases responses time (Guez & Naveh-Benjamin, 2006).

To summarize, contrary to my initial prediction, PF does not appear to stem from participants spending more time searching memory resulting in more correct responses. Instead, for correct responses, there is a marginal difference in the τ parameter estimates for A-Br and C-D pairs. This difference is larger and more reliable (as evidenced by non-overlap of the 95% CI) for incorrect responses indicating that A-Br pairs included more, faster responses relative to C-D pairs. However, participants are also taking longer to terminate their memory search for A-Br pairs relative to C-D pairs when no response is provided, as also evidenced by differences in τ. Thus participants seem to be spending more time searching memory for A-Br pairs, just not in the manner that I had initially predicted.

General Discussion

The general aim of the current project was to understand how memories for related material interact to our benefit and detriment. Specifically, I wanted to better understand the observation that memory for a pair items is sometimes improved if there are previous conflicting memories for the same material, a phenomena called proactive facilitation (PF). Three potential explanations for PF proposed by Aue et al. (2012) were considered. First, I examined whether participants were adopting a different response thresholds for either A-Br or C-D pairs. Second, I examined whether participants were encoding the A-Br pairs more completely during their presentation in list 2. Third, I examined whether participants were spending additional time searching memory for A-Br pairs. The data tended to be most consistent with a better, or at least differential, encoding explanation although the explanation is incomplete. Next, I discuss the limitations and potential modifications to the explanation that could more completely explain the constellation of observed data.

Explanations for proactive facilitation

In general the data appear to be most consistent with the better encoding account. At first blush the data from Aue et al. (2012) looks distinctly like a classic response bias data. Indeed, participants are outputting a greater number of responses overall, both correct and incorrect, but this is not the entire story. There are a number of problems for this account. Most notably, Experiments 1 and 2 were designed to eliminate a conservative or liberal response threshold, respectively, and there was no evidence that participants were changing the quality of response that they provided in either experiment. Moreover, if it were a global shift in response threshold that were driving the willingness to respond for both correct and incorrect responses, that would not be able to explain the data from Experiments 4 and 7 where PF is observed in the absence of PI. Lastly, if a liberal response threshold were being adopted one might expect it to be influenced by the familiarity of the cue at test. For instance, cue familiarity has been demonstrated to influence prospective memory judgments even when it dissociates from memory performance such as in interference paradigms (Eakin, 2005; Metcalfe et al., 1993; Schwartz & Metcalfe, 1992). The same memory judgments also can influence the duration of memory search. Nelson, Gerler, and Narens (1984) noted a positive correlation between a participants rated feeling of knowing for a piece of information and the latency for which they decided to terminate unsuccessful searches. Searches tended to go on longer when participants felt they knew the answer. Likewise, in Diller et al.'s (2001) model for cued recall the decision to terminate a memory search is influenced, in part, by the familiarity of the cue (see also Metcalfe, 1993). Thus, the fact that PF appears to be most prominent when both the cue and the target are repeated and not when simply the cue is repeated speaks against the longer memory search and differential response threshold hypotheses.

The better encoding account remains of the proposed hypotheses and is explicitly supported by Experiments 3, 4, and 6 in particular, and potentially Experiment 7. Experiments 3 and 4 manipulated the repetition of the cue and target independently during list 2 presentation. Critically, PF was most robust when both the cue and the target had been studied in list 1. In Experiment 6 the PF benefit for A-Br pairs persists across a long retention interval that would not be the case if participants were simply relying on cue familiarity to change their response threshold or search termination decisions. Lastly, for Experiment 7 the data from the ex-Gaussian modeling hinted at slightly faster responses for A-Br pairs that are consistent with a strength interpretation. The problem with a strength interpretation, as mentioned earlier, is that if it were simply the case that a more complete version of A-Br pairs is being stored then one would not expect to observe any PI. Again using Diller et al.'s (2001) cued recall framework as an example, under these circumstances at test the cues for A-Br pairs should match themselves better, and the influence of competitors would be attenuated. Likewise, targets would be recovered more easily thereby boosting performance. However, what is happening in the data appears to be two separate phenomena. If indeed better encoding is taking place, perhaps it is the case that it is not happening on every trial but rather a subset of trials. Recently, Wahlheim and Jacoby (2013) have posited a mechanism through which such a benefit could manifest.

Wahlheim and Jacoby (2013) have recently observed a form of PF that they suggest is driven by study phase retrieval. Their design is somewhat similar to the one employed in the current experiments. Participants study a list of weakly associated word pairs during an initial list. During a second list participants are presented with another list of pairs where the cue item is repeated but the target is replaced with a new, weakly associated target (i.e., A-B, A-D). During list 2, participants are told that some of the items from list 1 may be repeated, and they should try to detect the change and recall the list 1 target if a change is detected. At test participants are given a cue and asked to recall the word it was studied with on the most-recent list. Afterward, participants are asked about their phenomenological experience during recall. As discussed earlier, Wahlheim and Jacoby (Exp 1) find that participants were able to recall the list 2 target more often, relative to a C-D control, if the list 1 target came to mind first during free recall. Wahlheim and Jacoby have subsequently demonstrated similar effects in list discrimination (Jacoby Wahlheim, & Yonelinas, 2013), recency judgments (Jacoby & Wahlheim, 2013), and spacing effects (Wahlheim, Maddox, & Jacoby, 2014). The authors have explained their observation of PF using Hintzman's (2004, 2010) theory of recursive reminding for judgments of frequency and recency.

Hintzman (2004, 2010) proposed that when an item is encountered a second (or *n*th) time that spontaneous recall of the earlier events associated with the item in the experimental context could occur. This phenomenon is called study phase retrieval. The conditions of the earlier presentation are integrated into the memory for the current presentation in such a way that preserves item and order information, this is the recursive representation. As it relates to

Wahlheim and Jacoby (2013), they suggested that when a participant studies the pair A-B in list 1 and A-D in list 2, if the participant detected that A had been studied previously then the memory for the list 2 presentation would be A-D-B. Additionally, Burton et al. (2013) also found evidence of PF (they call it associative facilitation) and found that it was most robust when participant were instructed to form mediators between the A-B and A-D pairs. Benjamin & Tullis (2010) have used recursive reminding to explain distributed practice effect, the rationale being that when reminding takes place encoding of the second presentation is potentiated. Benjamin & Tullis provide a statistical model for distributed practice data based on these ideas. While the idea has been discussed with respect to associative data, no statistical or process models have been developed.

One parsimonious explanation may be that better encoding of the A-Br pairs takes place during list 2 as a consequence of study-phase retrieval. When a pair of items is presented for study, a participant might evaluate the familiarity of the presented item. If the items are sufficiently familiar then they are identified as having been studied before and the item's information from the existing (e.g., A-B) trace is updated and used in the list 2 trace thereby associating it with the new target (e.g., A-Br). The same process would take place for the target. The detection and updating of existing memory traces during study has been implemented in modeling the strength based mirror effect (Criss, 2006; Shiffrin & Steyvers, 1997) and has been demonstrated during testing (Criss, Malmberg, & Shffrin, 2011) as a component of an explanation of output interference. The idea that adding additional information (e.g., adding a new word or context to a trace) as opposed to updating existing information is novel. Such an explanation would also account for the fact that the current data appear to reflect a combination of PF and PI. Interference from other studied items would likely make the detection of A-Br pairs difficult and would not occur on every trial. Thus, memory performance for A-Br pairs could be better or worse depending on whether it was updated correctly during list 2. Wahlheim and Jacoby (2013) indirectly demonstrated such variability in repetition detection in A-D pairs.

Encoding during test and recall during encoding

Beyond the specific explanations for PF, the observation of PF has implications for memory research more generally. In the current paper I repeatedly observed PF in experimental designs that generate interference, but I also observed PF when interference was absent. Specifically, this was the case when interference was prevented by eliminating previous associations to material (Experiment 4) and by providing an experimental release from interference (Experiment 7). The persistence of PF in these scenarios indicates that PF is a phenomenon at least partially independent from PI. As such, the phenomena provide a novel perspective into the nature of memory.

Indeed, the current data reinforce the idea that memory is far more complicated than the outdated notion that learning only occurs during study and recall only occurs during test. There is ample evidence indicating that the encoding of information occurs while testing both to our detriment (e.g., output interference [Criss et al., 2011; Malmberg et al., 2012]) and to our benefit (e.g., retrieval-based learning [see Karpicke, Lehman, & Aue, 2014 for a recent review]). Additionally, the current explanation for PF put forth in the previous section suggests that memory retrieval is taking place during the encoding of information. Although this is not something that is measured in the current data, Wahlheim and Jacoby (2013) specifically asked participants whether they detected that an item had been studied previously. As discussed above, they found that when a participant noticed a change and the previous association (i.e., A-B) was accessible, they tended to recall the most recent association (i.e., A-D) more often. In other words, participants are learning while being tested and appear to be spontaneously testing themselves while studying information. Moreover, the fact that all of the participants in the current experiments (except Experiment 7) and those of Aue et al. (2012) were unaware that their memory was going to be tested indicates that this is something that occurs spontaneously and is not the result of a particular encoding strategy adopted by a subset of participants.

The benefit extends beyond the confines of memory experimentation. Indeed, the idea that participants are building on existing memories to aid in learning new information is similar to a well-known learning strategy encouraged by educators called self-explanation. Berry (1983) found that children were more accurate at certain types of problem solving if they discussed the problem and related to their existing knowledge while they were attempting to solve it. Similarly, participants may be drawing existing representation of the studied information to provide a starting point for the encoding of the new pairs in list 2.

Future directions and conclusion

The current data also raised new questions about PF. Perhaps the biggest question is what is it exactly that is contained in the memory representation for A-Br pairs in list 2. Wahlheim and Jacoby (2013) employed an A-B, A-D design for their experiment that is manipulating the repetition of the cue as we did in Experiment 3. This means that the representation for an A-B pair following list 2 would be A-D-B, if the repetition were noticed. In the current experiment, I employed an A-B, A-Br design where *both* the cue and the target are repeated across lists. Under the recursive representation logic for A-Br pairs the representation for list 2 studied pairs would contain the list 1 partners for both items of the A-Br pair if the repetition were noticed. For example, as is depicted in Table 2 if a participant studied the pairs *Absence-Hollow* and *Pupil-River* during list 1 that are then rearranged for list 2 so that they study *Absence-River* then the extension of the Wahlheim & Jacoby explanation is that study of *Absence-River* (assuming noticing of the repetition) would be encoded as *Hollow-Absence- Pupil-River*.

While it is impossible to examine the contents of the member trace directly, one could test a participant's memory for a remote association between the items that were not studied together but were studied *with* items that were studied together. An interesting test of this idea may be to give a participant *Pupil* and measure how often they respond with *Absence*. The two items were never studied together but they would be associated in a recursive representation. In this design, *Pupil* is an independent cue because the two words were never studied together, but would be directly associated if stored in a single representation. This is type of response would be similar in concept to the remote intrusions described by Provyn, Sliwinski, and Howard (2007).

A second, and perhaps more problematic issue, is explaining the variability that is producing the combination of PF and PI that is observed in the current data. It is unclear what exactly is happening during list 2 study that is leading to only some of the pairs being better/differentially encoded while others are not. Wahlheim and Jacoby (2013) describe the participant's phenomenological experience during the final test, but not necessarily what is happening during list 2 study. If it is indeed study-phase retrieval that it triggering the better/differential encoding of pairs, why are some pairs retrieved and not others? Additionally, can that same explanation also account for why PF has gone largely unnoticed in the literature for so very long? As discussed Burton et al. (2013) have suggested that the differences in paired associates techniques (e.g., learning to a criterion) may have contributed to masking the PF effects.

The current project focused on understanding how memories for recent experiences interacted with recent memories for related information. I examined this by having participants attempt to form new memories that conflicted with a recent memory by having a different pairing of the material (e.g., a new word with an old face). The result of the experiments was the demonstration the phenomena of proactive facilitation across a variety of experimental scenarios. During these experiments, I was able to rule out decisional factors such as an altered response threshold. The current data appear to be consistent with the idea that PF is instead a feature of memory that has been largely overlooked. I suggest that when some items are encountered a second time in a given study context that better versions of these items are encoded and participants may try harder before giving up the search of memory. This facilitates recall of these pairs relative to a control pair that had not been encountered previously.

Appendix A: First keystroke and response duration reaction time data

First keystroke RT

The response distributions for First keystroke (KS) RT are presented in Figure A1. The differences in the tails for correct responses and no recalls that were observed for Signal RT are not present in these data. However there is still a tendency for C-D pairs to have slightly longer incorrect responses relative to A-Br pairs.

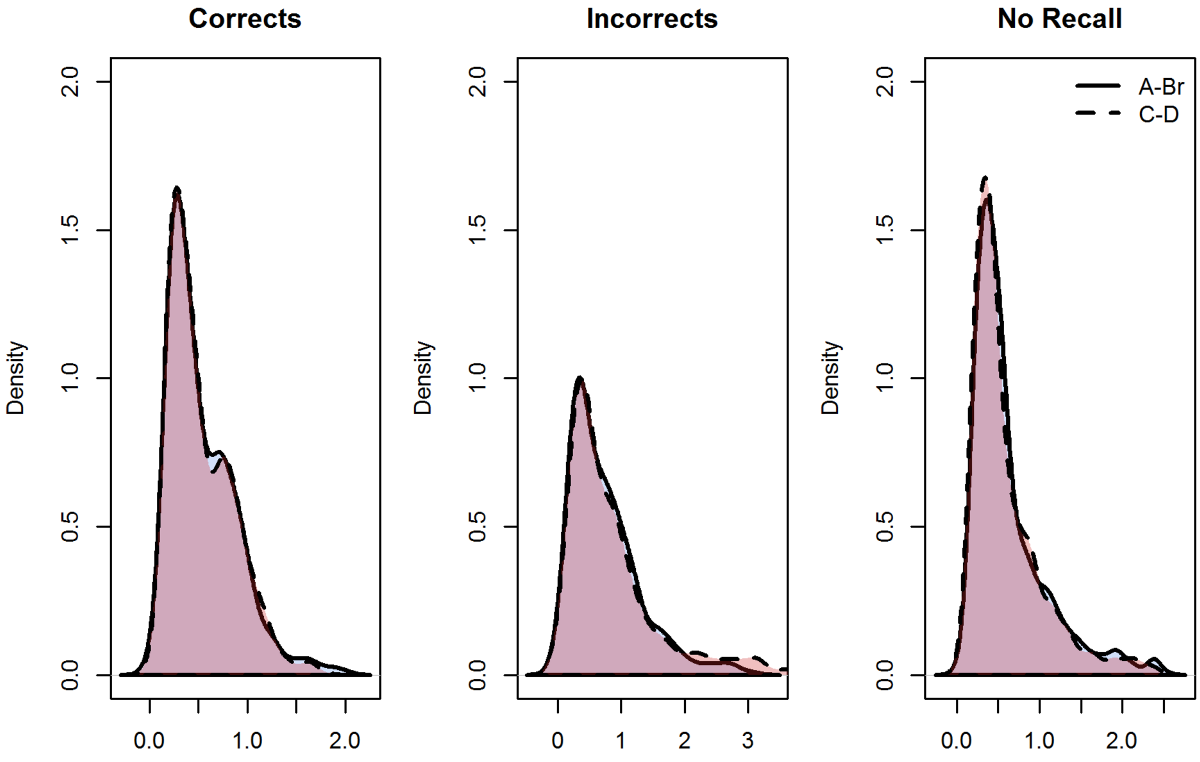

Figure A 1. Response distributions for First keystroke RT for A-Br and C-D responses.

As can be seen in Figure A2 the ex-Gaussian distribution tended to capture the general pattern of the First KS RT quite well. The best fitting parameter values as well as descriptive statistics are provided in Table A1.

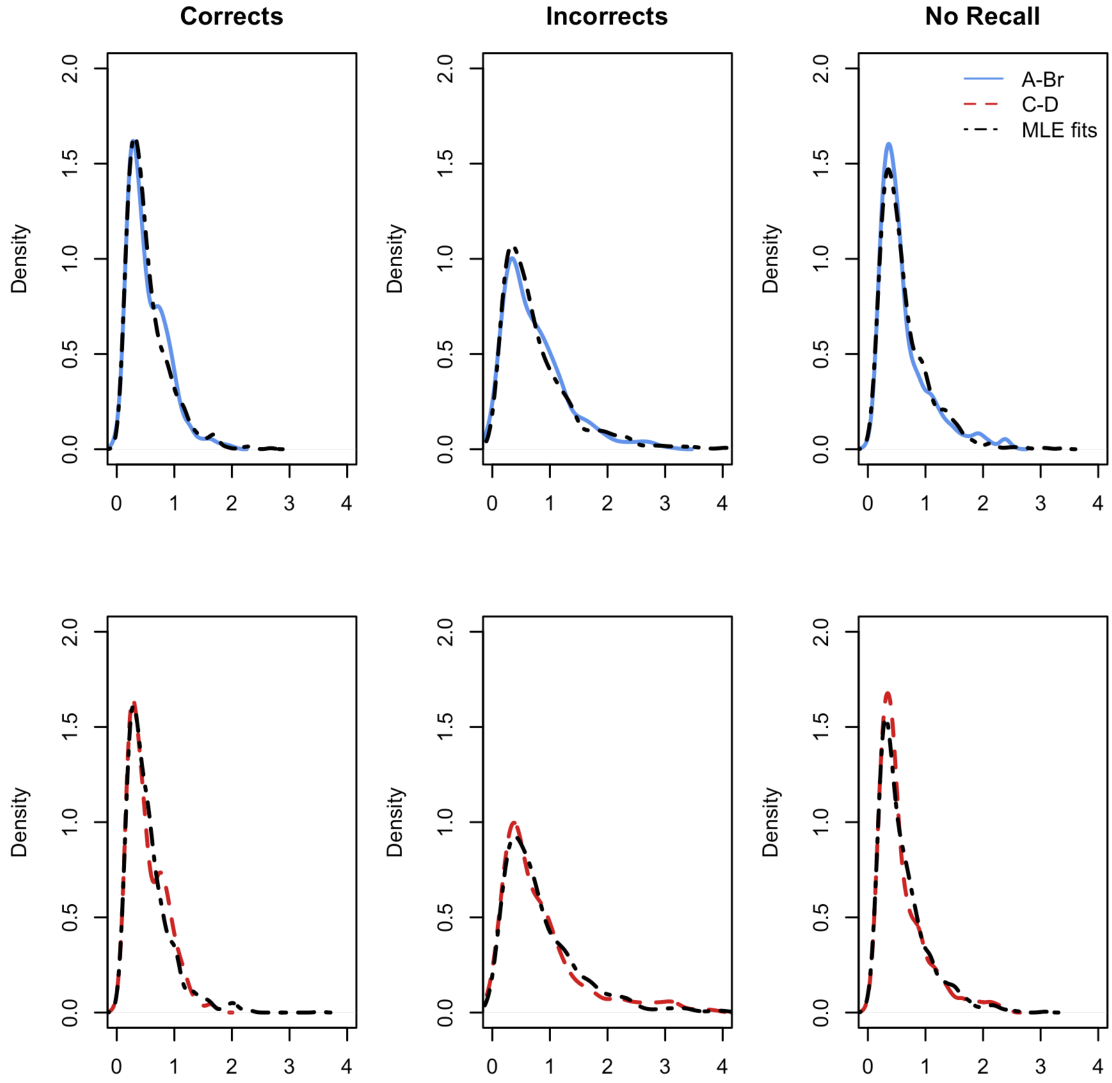

Figure A 2. Model fits of an ex-Gaussian distribution to the First KS RT data.

Memory performance for A-Br and C-D pairs are plotted as a function of First KS RT quantile in Figure A3. There are no clear patterns to the data. A-Br and C-D quantile pairs tend to occupy the same vertical space along the y-axis indicating no difference in reaction times.

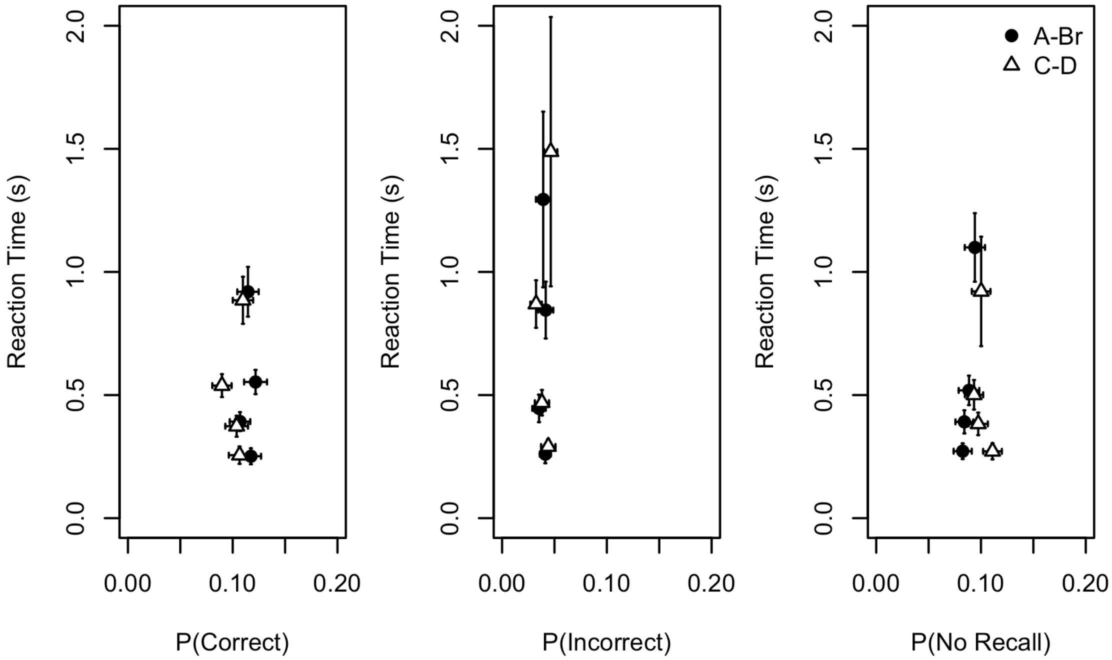

Figure A 3. Memory performance plotted as a function of First KS RT quantile.

Response Duration

The distributions for Response Duration are presented in Figure A4. There is substantial overlap between the distributions for A-Br and C-D pairs, although for correct responses there is again a trend toward response taking slightly longer, in this case to complete their response, for C-D pairs. In the ex-Gaussian modeling this manifests as C-D pairs tending to have a larger τ than the A-Br pairs, as was the case with the Signal RT data. The ex-Guassian model fits to the data are provided in Figure A5 and the descriptives for the distributions and the best fitting ex-Gaussian model parameters are provided in Table A2. Memory performance for A-Br and C-D pairs are plotted as a function of Response Duration quantile in Figure A6. Again, there are no clear patterns to the data. A-Br and C-D quantile pairs tend to occupy the same vertical space along the y-axis indicating no difference in reaction times.

Table A 1. Best fitting ex-Gaussian parameters for the first keystroke RT

	Behavioral data Descriptives		Ex-Gaussian Model fits to behavioral data		
	M	*SE*	μ	σ	τ
Correct					
A-Br	.604	.051	.166 (.156, .182)	.076 (.062, .091)	.372 (.348, .392)
C-D	.655	.065	.167 (.155, .183)	.067 (.054, .083)	.362 (.338 .382)
Incorrect					
A-Br	.707	.053	.162 (.134, .199)	.103 (.072, .135)	.586 (.524,.644)
C-D	.769	.054	.160 (.130, .189)	.098 (.069, .123)	.686 (613, .765)
No Recall					
A-Br	.585	.038	.197 (.183, .214)	.072 (.055, .087)	.430 (.394, .462)
C-D	.558	.036	.184 (.169, .196)	.070 (.056, .081)	.411 (.383, .441)

Note. 95% confidence intervals generated based on 10,000 bootstrapped samples.

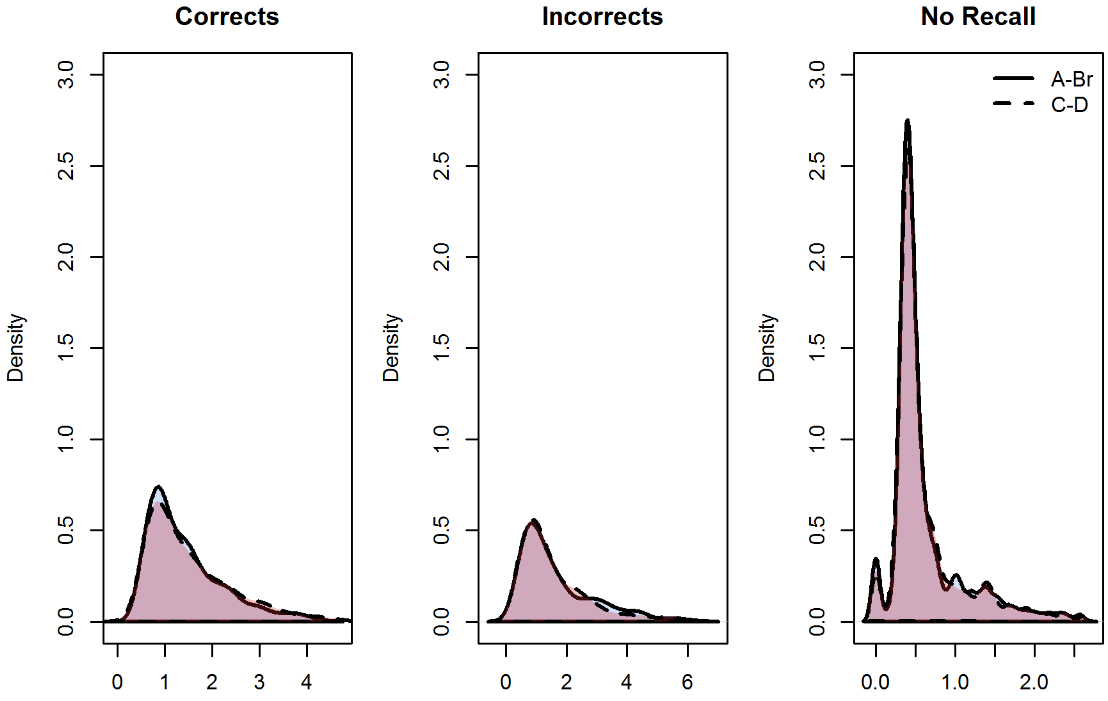

Figure A 4. Response distributions for Response Duration for A-Br and C-D responses.

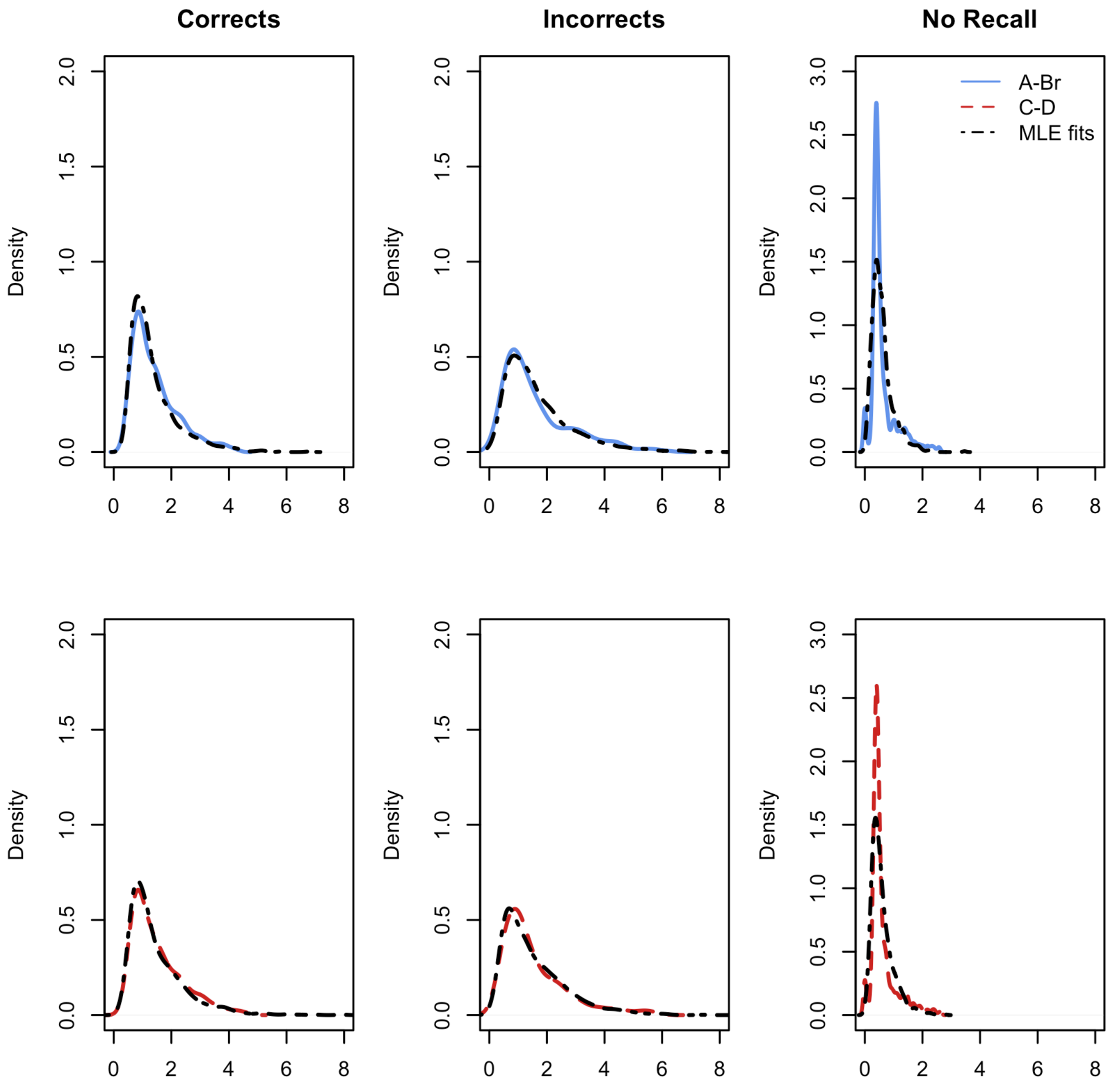

Figure A 5. Model fits of an ex-Gaussian distribution to the First KS RT data

Table A 2. Best fitting ex-Gaussian parameters for the response duration

	Behavioral data Descriptives		Ex-Gaussian Model fits to behavioral data		
	M	*SE*	μ	σ	τ
Correct					
A-Br	1.37	.069	.564 (.528, .595)	.128 (.101, .151)	.859 (.808, .914)
C-D	1.32	.056	.547 (.519, .581)	.129 (.107, .156)	.958 (.898, 1.01)
Incorrect					
A-Br	1.60	.103	.418 (.360, .478)	.140 (.100, .177)	1.23 (1.10, 1.36)
C-D	1.59	.091	.354 (.322, .479)	.071 (.021, .176)	1.20 (1.05, 1.32)
No Recall					
A-Br	.573	.045	.237 (.216, .254)	.130 (.113, .141)	.380 (.345, .425)
C-D	.568	.047	.245 (.227, .259)	.123 (.109, .135)	.369 (.337, .407)

Note. 95% confidence intervals generated based on 10,000 bootstrapped samples.

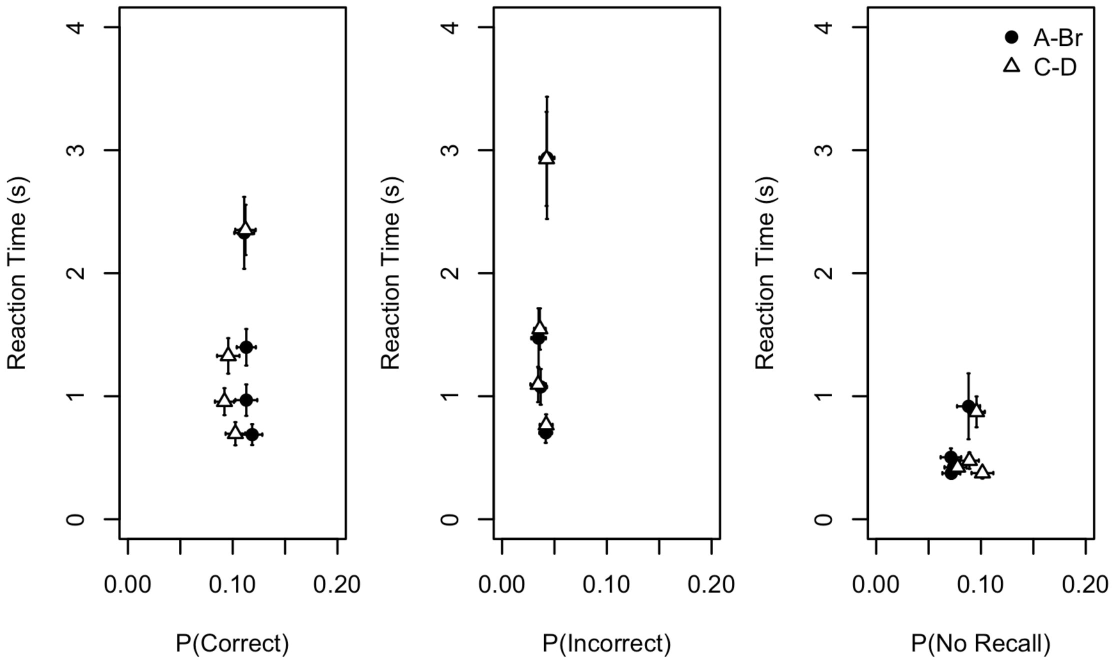

Figure A 6. Memory performance plotted as a function of Response Duration quantile

References

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. Bjork, R. A. Bjork, E. Bjork, R. A. Bjork (Eds.), *Memory* (pp. 237-313). San Diego, CA US: Academic Press. http://dx.doi.org/10.1016/B978-012102570-0/50010-0

Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal Of Memory And Language*, *66*(1), 109-122. http://dx.doi.org/10.1016/B978-012102570-0/50010-0

Barnes, J. M., & Underwood, B. J. (1959). 'Fate' of first-list associations in transfer theory. *Journal of Experimental Psychology, 58(2)*, 97-105. http://dx.doi.org/10.1016/B978-012102570-0/50010-0

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37(3)*, 379-384.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445-459. http://dx.doi.org/10.1016/B978-012102570-0/50010-0

Benjamin, A. S. & Tullis, J. G. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*(3), 228-247.

Berry, D. C. (1983). Metacognitive experience and transfer of logical reasoning. *Quarterly Journal of Experimental Psychology*, *35*, 39-49.

Bruce, D., & Weaver, G. E. (1973). Retroactive facilitation in short-term retention of minimally learned paired associates. *Journal of Experimental Psychology*, *100*(1), 9-17. http://dx.doi.org/10.1037/h0035488

Burns, D. J. (1989). Proactive interference: An individual-item versus relational processing account. *Journal Of Memory And Language*, *28*(3), 345-359. http://dx.doi.org/10.1016/0749-596X(89)90038-7

Burton, R. L., Lek, I. & Caplan, J. B. (*2013*). *Associative independence revisited: Evidence for associative facilitation and competition in AB/AC learning.* Manuscript in preparation.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*(4), 461-478.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal Of Memory And Language*, *64*(2), 119-132. http://dx.doi.org/10.1016/j.jml.2010.10.001

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31(6)*, 1199-1212. http://dx.doi.org/10.1037/0278-7393.31.6.1199

Criss, A. H., Malmberg, K. J., & Shiffrin, R.M. (2011). Output Interference in Recognition Memory. *Journal of Memory and Language, 64*(4), 316-326.

Crowder, R. G. (1976). *Principles of learning and memory*. Oxford England: Lawrence Erlbaum.

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27(2),* 414-435. http://dx.doi.org/10.1037/0278-7393.31.6.1199

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452-478. http://dx.doi.org/10.1037/0033-295X.108.2.452

Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of memory and language, 52,* 526-534.

Guez, J., & Naveh-Benjamin, M. (2006). Divided attention at encoding and retrieval for once- and thrice-presented items: A micro-level analysis of attentional costs. *European Journal Of Cognitive Psychology, 18(6)*, 874-898. doi:10.1080/09541440500485854

Hanczakowski, M., Pasek, T., Zawadzka, K., & Mazzoni, G. (2013). Cue familiarity and 'don't know' responding in episodic memory tasks. *Journal Of Memory And Language, 69,* 368-383. http://dx.doi.org/10.1016/j.jml.2013.04.005

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, *32*(2), 336-350.

Hirshman, E., Burns, D. J., & Kuo, T. (1993). Examining a processing tradeoff explanation of proactive interference. *Memory & Cognition*, *21*(1), 5-10. http://dx.doi.org/10.3758/BF03211159

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208-233. doi:10.1037/0033-295X.96.2.208

Jacoby, L. L. & Wahlheim, C. N. (2013). On the important of looking back: The role of recursive reminding in recency judgements and cued recall. *Memory and Cognition, 41,* 625-637.

Jacoby, L. L., Wahlheim, C. N., & Yonelinas, A. P. (2013). The role of detection and recollection of change in list discrimination. *Memory and Cognition, 41,* 638-649.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 112.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of Learning and Motivation, 61,* (p. 237-284). San Diego, CA: Elsevier Academic Press.

Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting Context in Context. In Narine, J. S. (Ed), *The Foundations of Remembering: Essays in honor of Henry L. Roediger, III.,* (pp. 171-189). New York, NY: Psychology Press.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal Of Experimental Psychology: General*, *123*(3), 297-315. http://dx.doi.org/10.1037/0096-3445.123.3.297

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103(3),* 490-517. http://dx.doi.org/10.1037/0033-295X.103.3.490

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

Lawrence, M. A. (2013). ez: Easy analysis and visualization of factorial experiments.. R package version 4.2-2. http://CRAN.R-project.org/package=ez

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review, 120(1)*, 155-189. doi: 10.1037/a0030851

Light, L. L., Patterson, M. M., Chung, C., & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition, 32(7)*, 1182-1193. doi:10.3758/BF03196891

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers*, *28*(2), 203-208. http://dx.doi.org/10.3758/BF03204766

Macmillan, N. A. (1993). Signal detection theory as data analysis method and psychological decision model. In G. Keren & C. Lewis (Eds.), *Methodological and Quantitative Issues in the Analysis of Psychological Data: Vol 1* (pp. 21-57). Hillsdale, NJ: Erlbaum.

Malmberg, K. J., Criss, A. H., Gangwani, T. & Shiffrin, R. M. (2012). Overcoming the Negative Consequences of Interference that Results from Recognition Memory Testing. *Psychological Science*, 23, 115-119.

Malmberg, K. J., & Shiffrin, R. M. (2005). The 'One-Shot' Hypothesis for Context Storage. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *31*(2), 322-336. http://dx.doi.org/10.1037/0278-7393.31.2.322

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68-78.

Massidda, D. (2013). retimes: Reaction Time Analysis. R package version 0.1-2. http://CRAN.R-project.org/package=retimes

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: a subjective-likelihood approach to the effects of experience in recognition memory. *Psychological review*, *105*(4), 724-760.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Mmeory, and Cognition, 19(4),* 851-861

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89(6),* 609-626. doi:10.1037/0033-295X.89.6.609

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282–300.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal Of Experimental Psychology: Learning, Memory, And Cognition, 27(2)*, 384-413. doi:10.1037/0278-7393.27.2.384

Postman, L., & Stark, K. (1964). Proactive inhibition as a function of the conditions of transfer. *Journal Of Verbal Learning & Verbal Behavior, 3(3)*, 249-259. http://dx.doi.org/10.1016/S0022-5371(64)80050-5

Postman, L., Stark, K., & Burns, S. (1974). Proactive inhibition on unpaced tests of retention. *The American Journal of Psychology, 87(1/2)*, 33-56. http://dx.doi.org/10.2307/1421998

Postman, L. & Gray, W. (1977). Maintenance of prior associations and proactive inhibition. *Journal of Experimental Psychology: Human Learning and Memory, 3(3)*, 255-263. http://dx.doi.org/10.1037/0278-7393.3.3.255

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology*, *79*(1, Pt.1), 168-177. http://dx.doi.org/10.1037/h0026932

Postman, L. & Underwood, B. (1973). Critical issues in interference theory. *Memory and Cognition, 1*, 19-40. http://dx.doi.org/10.3758/BF03198064

Provyn, J. P., Sliwinski, M. J., & Howard, M. W. (2007). Effect of age on contextually mediate associations in paired associate learning. *Psychology & Aging, 22(4)*, 846-857.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Robbins, D., & Bray, J. F. (1974). The spacing effect and the AB, AC paradigm: Evidence for retroactive facilitation. *Journal of Experimental Psychology*, *103*(3), 420-425.

Roediger, H. L., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, *13*(1), 1-7. http://dx.doi.org/10.3758/BF03198437

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievabiliby enhances feeling of knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(5),* 1074-1083.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM - retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4 (2)*, 145-166. http://dx.doi.org/10.3758/BF03209391

Tulving, E. (1985). *Elements of Episodic Memory.* Oxford University Press: Oxford.

Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal Of Verbal Learning & Verbal Behavior*, *13*(2), 181-193. http://dx.doi.org/10.1016/S0022-5371(74)80043-5

Underwood, B. J. (1949). Proactive inhibition as a function of time and degree of prior learning. *Journal Of Experimental Psychology*, *39*(1), 24-34. doi:10.1037/h0059550

Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76(6)*, 559-573. http://dx.doi.org/10.1037/h0028143

Verde, M. F., & Rotello, C. M. (2004). Strong memories obscure weak memories in associative recognition. *Psychonomic Bulletin & Review, 11(6)*, 1062-1066. doi:10.3758/BF03196737

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41(1),* 1-15. http://dx.doi.org/10.3758/s13421-012-0246-9

Wahlheim, C. N., Maddox, G. B., Jacoby, L. L. (2014). The role of reminding in the effects of spaced reptitions on cued recall: Sufficient but not necessary. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40(1),* 94-105.

Curriculum Vitae

WILLIAM R. AUE, M.A.

Syracuse University

Department of Psychology

430 Huntington Hall

Syracuse, New York 13244

(315) 842-4283 william.aue@gmail.com

https://sites.google.com/site/williamaue/

Education

Purdue University, West Lafayette, Indiana

Post-doctoral research associate, Department of Psychological Sciences, August 2014

Advisor: Jeffrey D. Karpicke, Ph.D.

Syracuse University, Syracuse, New York

Ph.D., Experimental Psychology, 2014

Dissertation: "Understanding Proactive facilitation in cued recall"

Advisor: Amy H. Criss, Ph.D.

University of West Florida, Pensacola, Florida

M.A., Psychology, 2008

Thesis: "Periodicities in performance of visual continuous performance tasks"

Advisor: James E. Arruda, Ph.D.

Butler University, Indianapolis, Indiana

B.A., Psychology, 2004

Research Interests

Human episodic memory, Computational modeling of memory, Associations in memory, Sustained attention, Experimental methods

Manuscripts in Preparation

Aue, W. R., Fontaine, J. M., & Criss, A. H. (in preparation). Context variability and memory for items and associations.

Aue, W. R., Criss, A. H., & Novak, M. (in preparation). Understanding proactive facilitation of memory for word-face pairs

Publications

Criss, A. H., Salomão, CR., Malmberg, K. J., Aue, W. R., Kılıç, A., & Claridge, M. (under review). Release from output interference: Words and faces.

Aue, W. R., Criss, A. H., & Prince. M. A. (in revision). Dynamic memory searches: Output interference selectively impacts episodic memory.

Howard, M. H., Shankar, K. H., Aue, W. R., and Criss, A. H. (in press). A quantitative model of internal time in memory. *Psychological Review.*

Criss, A. H., Aue, W. R., and Kılıç, A. (2014). Age and response bias: Evidence from the strength based mirror effect. *The Quarterly Journal of Experimental Psychology, 67(10),* 1910-1924.

Karpicke, J. D., Lehman, M., Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *The Psychology of Learning and Motivation*, *61,* 237-284.

Aue, W. R., Criss, A. H., Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal of Memory and Language, 66(1)*, 109-122.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language, 64,* 119-132.

Aue, W. R., Arruda, J. E., Kass, S. J., & Stanny, C. J. (2009). Cyclic variations in sustained human performance. *Brain & Cognition, 71,* 336-344.

Julian, M., Bohannon, J. N., & Aue, W. R. (2009). Measures of flashbulb memories: Are elaborate memories consistently accurate? In O. Luminet & A. Curci (Eds). *Flashbulb Memories: New Issues and New Perspectives* (pp. 99-122). New York, NY: Psychology Press

Arruda, J. E., Zhang, H., Coburn, K. L., Amoss, R. T., & Aue, W. R. (2009). Rhythmic oscillations in quantitative EEG measured during a continuous performance task. *Journal of Applied Psychophysiology and Biofeedback, 34*, 7-16.

Presentations

Aue, W. R., Prince, M., & Criss, A. H. (2013, November). Examining output interference in memory for factual information. Poster accepted for presentation at the annual meeting of the Psychonomic Society. Toronto, Canada.

Wilson, J. H., Aue, W. R., & Criss, A. H. (2013, November). The effect of cue and target strength in cued recall. Poster presented at the annual meeting of the Psychonomic Society. Toronto, Canada.

Aue, W. R., Hemmer, P., Criss, A. H. (2013, July). A Bayesian Analysis of Bias in Single Item Recognition for Continuous Word frequency. Poster presented at the annual meeting of the Society for Mathematical Psychology. Potsdam, Germany.

Aue, W. R., Hemmer, P. & Criss, A. H. (2013). A bayesian analysis of bias in single item recognition for continuous word frequency [Abstract]. In M. Knauff, M. Pauen, N.

Sebanz, & I. Wachsmuth (Eds.) *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (p. 3873). Austin, TX:

Cognitive Science Society.

Aue, W. R., Fontaine, J., and Criss, A. H. (2013, May). Word frequency and context variability effects dissociate as a function of test type. Poster presented at the annual meeting of the Context and Episodic Memory Symposium. Philadelphia, PA.

Hemmer, P., Aue, W. R. & Criss, A. H. (2013, May). The relationship between multiple memory tasks. Talk given at the annual meeting of the Context and Episodic Memory Symposium. Philadelphia, PA.

Salomão, C., Claridge, M., Kılıç, A., Aue, W. R. & Criss, A. H. (2013, March). Release from Output Interference in Long Term Memory. Poster presented at the annual meeting of the Eastern Psychological Association. New York, NY.

Fontaine, J. M., Holder, J. N., Kılıç, A., Aue, W. R. & Criss, A. H. (2013, March). Examining Output Interference as a Function of Study List Length. Poster presented at the annual meeting of the Eastern Psychological Association. New York, NY.

Aue, W. R., Fontaine, J., and Criss, A. H. (2012, November). Context variability and memory for items and associations. Poster presented at the annual meeting of the Psychonomic Society. Minneapolis, MN.

Aue, W. R. and Criss A. H. (2012, July). Item strength and associative information: Examining the role of item repetition on subsequent cued recall. Poster presented at the annual meeting of the Society of Mathematical Psychology. Columbus, OH.

Aue, W. R. and Criss A. H. (2012, May). Item strength and associative information: Examining the role of item repetition on subsequent cued recall. Poster presented at the annual Context and Episodic Memory Symposium. Bloomington, IN.

Fontaine, J., Aue, W. R., & Criss, A. H. (2012, March). Examining the effects of word frequency and context variability on episodic memory. Poster presented at the annual meeting of the Eastern Psychological Association. Pittsburgh, PA.

Aue, W. R. & Criss, A. H. (2011, November). Items strength benefits associative encoding. Poster presented at the annual meeting of the Psychonomic Society. Seattle, WA.

Aue, W. R. & Criss, A. H. (2011, July). The role of cue strength in cued recall. Poster presented at the annual meeting for the Society for Mathematical Psychology. Somerville, MA.

Aue, W. R., Criss, A. H., & Fischetti, N. W. (2011, May). Associative information in memory: Evidence from cued recall. Poster presented at the annual Context and Episodic Memory Symposium. Philadelphia, PA.

Aue, W. R., Criss, A. H., & Fischetti, N. W. (2010, November). Item repetition within (but not across) pair type helps and harms cued recall. Poster presented at the annual meeting of the Psychonomic Society. St. Louis, MO.

Criss, A.H., & Aue, W.R. (2010, August). Using cued recall to move beyond task specific models. Models of Episodic Memory Symposium (Organizer: Simon Dennis) at the Society for Mathematical Psychology Annual Meeting. Portland, OR.

Aue, W. R., Criss, A. H., Fischetti, N. W. (2010, July). Item repetition within (but not across) pair type helps and harms cued recall. Talk given at the Annual Summer Interdisciplinary Conference. Bend, OR.

Aue, W. R. & Criss. A. H. (2010, July). Investigating the representation of associative information. Poster given at the European Society for Cognitive Psychology summer school. Mallnitz, Austria.

Criss, A.H., Aue, W. R., & Smith, L. (2010, April). The Effects of Word Frequency and Context Variability in Cued Recall. Context in Episodic Memory Symposium, Philadelphia, PA.

Aue, W. R., Arruda, J. E., Kass, S. J., & Stanny, C. J. (2008, May). Periodicities in performance on visual continuous performance tasks. Poster presented at the annual meeting of the Association for Psychological Science. Chicago, IL.

Arruda, J. E. & Aue, W. R. (2008, May). Cyclic variations in sustained human performance. Presentation at the Department of Defense Human Factors Engineering Technical Avisory Group Meeting #59. Eglin AFB, Destin, FL

Aue, W. R., Qualls, E. K., & Arruda, J. E. (2007, Feb). Latent structure of the Visual Analog Mood Scales. Poster presented at the annual meeting of the Southeastern Psychological Association. New Orleans, LA.

Qualls, E. K., Aue, W. R., Sconfietti, L., & Arruda, J. E. (Feb 2007). The effect of sleep on the reliability of the VAMS. Poster presented at the annual meeting of the Southeastern Psychological Association. New Orleans, LA.

Julian, M., Aue, W. R., & Bohannon, J. N. (2006, May). The effect of exposure on quantity of recall in flashbulb memory. Poster presented at the 18th Annual meeting of the Association for Psychological Science. New York, NY.

Aue. W. R. & Arruda, J. E. (2006, April). Perceptual bias and scales of measurement associated with the Visual Analog Scales. Paper presented at the 3rd Annual University of West Florida Sea Stars Science and Technology Research Symposium. Pensacola, FL.

Aue. W. R. & Arruda, J. E. (2006, March). Perceptual bias and scales of measurement associated with the Visual Analog Scales. Poster presented at the annual meeting of the Southeastern Psychological Association. Atlanta, GA.

Walch, S. E., Stanny, C. J., Mathews, S. R., Arruda J. E., McGee, H. A., & Aue, W. R. (2006, March). Enhancing the mentoring relationship. Conversation hour at the annual meeting of The Southeastern Psychological Association. Atlanta, GA.

Julian, M., Bohannon, J. N., & Aue, W. R. (2006, March). Measures of flashbulb memory: Are elaborate memories consistently accurate? Poster presented at the annual meeting of the Southeastern Psychological Association. Atlanta, GA.

Julian, M., Bohannon, J. N., & Aue. W. R. (2006, July). Measures of flashbulb memories: Are elaborate memories consistently accurate? Paper presented at the meeting of the International Society for Research on Emotion. Bari, Italy.

Aue, W. R., Kozlowski, A., Lesmeister, L., Green, S., & Bohannon, J. N. (2004, May). Confidence and consistency characterize flashbulb memories for the Columbia disaster. Poster presented at the sixteenth annual meeting of the Association for Psychological Science. Chicago, IL.

Aue, W. R., Rush, A., & Bohannon, J. N. (2004, March). Memory context effect in flashbulb memory: Columbia, Challenger, and 9/11. Poster presented at the meeting of the Southeastern Psychological Association. Atlanta, GA.

Aue, W. R., Kozlowski, A. Bohannon, J. N., Lesmeister, L., Matthews, Z., & Barga, S. (2004, March). Flashbulb memories are more consistent than everyday memories. Poster presented at the meeting of the Southeastern Psychological Association. Atlanta, GA.

Tyson, E., Barga, S., Aue, W. R., Kozlowski, A., & Matthews, Z. (2003, May). Flashbulb memories of the world trade center: Same day assessment with 3-month and 1 year follow-ups. Poster presented at the fifteenth annual meeting of the Association for Psychological Science. Atlanta, GA.

Aue, W. R., Rush, A., Ryan, H., & Bohannon, J. N. (2003, March). The effects of memory cues at recall on flashbulb memories. Poster presented at the meeting of the Southeastern Psychological Association. New Orleans, LA.

Grants and Awards

Society for Mathematical Psychology
Award: Student Travel Award
Amount: $200
Date: July 2013

Society for Mathematical Psychology
Award: Student Travel Award
Amount: $160
Date: July 2012

Society for Mathematical Psychology
Award: Student Travel Award
Amount: $200
Date: July 2011

Syracuse University, Department of Psychology
Eric F. Gardner Research Fellowship
Date: Fall 2010 – Spring 2011

European Society for Cognitive Psychology
Award: Admission to “Computational and Mathematical Modeling of Cognition” summer school (competitive)
Date: July 2010

Scholars of Engineering, Applied Sciences & Technology Annual Research Symposium, University of West Florida
Award: Departmental honors
Project title: Periodicities in performance on visual sustained attention tasks.
Date: April 2008

Scholarly and Creative Activities Committee, University of West Florida
Project title: Periodicities in performance on visual sustained attention tasks.
Amount: $525
Date: October 2007

Scholarly and Creative Activities Committee, University of West Florida
Project title: Perceptual Bias and Scales of Measurement Associated with the Visual Analog Scales.

Amount: $500 for travel

Date: November 2005

Research Experience

Graduate student, Memory Modeling Laboratory, Amy H. Criss, Ph.D., Fall 2008 to present

Masters Thesis, “Periodicities in performance of visual continuous performance tasks,” Spring 2007 to Spring 2008.

Program Analyst 2, Naval Aerospace Medical Research Laboratory, Benton Lawson, Ph.D., Spring 2007 to Summer 2008

Lab Coordinator, Autobiographical Memory Lab, Butler University, John N. Bohannon, Ph.D., Fall 2003 to Spring 2004

Teaching Experience

Courses taught

Cognitive Psychology

Introduction to Psychology

Laboratory in Cognitive Psychology

Neuroscience and Society

Affiliations

Cognitive Science Society

Association for Psychological Science

Society for Mathematical Psychology

Psychonomic Society