6-4-2019

Forget Me Not: Are Stronger Memories More Susceptible to Retrieval-Induced Forgetting?

Laura Lee Heisick
Louisiana State University and Agricultural and Mechanical College, lheisi1@lsu.edu

FORGET ME NOT: ARE STRONGER MEMORIES MORE SUSCEPTIBLE TO RETRIEVAL-INDUCED FORGETTING?

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
Laura Lee Heisick
B.S. Louisiana State University, 2014
M.A., Louisiana State University, 2017
August 2019

To Christina, for too many reasons to list.

Acknowledgments

Throughout the writing of this dissertation, I have received a great deal of support and assistance. I would first like to thank my advisor, Dr. Megan Papesh, whose feedback at every stage of this project (and every other project!) has been invaluable. To Megan – for your advice, patience, commitment to the research process, and availability, thank you. I can't say enough how much I appreciate your time and effort spent mentoring me, and your willingness to help whenever and however needed. Thank you for being in my corner, giving me advice, and tolerating me stopping by to chit chat so, so many times over the many years. I hope I don't become a time vampire in years to come!

To my committee members Dr. Jason Hicks, Dr. Melissa Beck, and Dr. Joseph Legoria – I am very grateful for your time and energy and willingness to trek all across campus for our meetings. I appreciate your thoughtful feedback and insight into my project, and this document is certainly better for it!

To Dr. Mike Hout – thank you for allowing me to use your program, and for your time and effort spent helping me modify it for this project. While wading through unfamiliar code was an undertaking on its own, this project couldn't have been completed without you! To Dynna Washington, Torrance Thomas, Niels Dickson, and Alexa Rabie – thank you for your assistance with data collection, and your patience with my extremely and overbearingly specific participant assignments. I certainly could not have completed this research in a timely manner without all of your help.

To Juan Guevara Pinto – I can't write enough about how grateful I am that we began the program at the same time. Having someone to complain with and at (and, okay, sometimes about), to bounce ideas off of, and who sympathizes with every step of the process has been so helpful. I'll miss our extremely productive Friday afternoon lab meetings.

To Jessica Hacker and Daniella Cash – I am so happy that you were both a part of my final year in the lab! Being able to turn around and chat or complain, to share exciting news or feelings of being overwhelmed, to commiserate over timelines and milestones – I am very grateful for our friendships.

To my parents, Doug and Glenda, and my sister, Kaylee, for supporting me throughout my many, many years of schooling. Thank you for all your help, and the many ways it was offered. You all have stood by me no matter what the circumstances, which I appreciate – and look, I *did* finish school!

And finally, to Christina. Without your help and support, I doubt I would have been able to complete this project – or any project. Thank you for believing in me, for encouraging me, and for supporting me. Enough said. I've never been good at this sort of thing, have I?

Table of Contents

Abstract

Successfully retrieving information sometimes causes forgetting of related, but unpracticed, information, termed retrieval-induced forgetting (RIF). One explanatory mechanism of RIF suggests related, but currently irrelevant, information is inhibited during retrieval, resulting in poorer memory for competing representations. Critically, this perspective suggests stronger memories are more susceptible to RIF because stronger representations produce additional competition when unpracticed. To resolve this competition, strong competing items are inhibited, resulting in the counterintuitive prediction that stronger memories are more likely to be forgotten. The aim of the current experiments was to replicate and extend recent work suggesting non-typical objects and own-race faces, both of which are associated with stronger memory traces, are more likely to be forgotten. In Experiment 1, participants studied and practiced typical and non-typical objects before memory was assessed through recognition or measures of perceptual similarity. Results showed object memorability influenced the magnitude of RIF: Non-typical (i.e., highly memorable) objects were more likely to be forgotten than typical (i.e., non-memorable) objects. However, RIF did not correspond with changes in perceived similarity. In Experiment 2, participants studied and practiced own- and other-race faces before memory was assessed, again through either recognition or similarity measures. Experiment 2 revealed no RIF for own- or other-race faces, and no corresponding changes in perceived similarity. These findings suggest that if memory traces are too weak to produce competition, no RIF is observed. Considered together, these results support inhibitory accounts of RIF, and suggest stronger memories produce additional competition that makes them more susceptible to forgetting.

Introduction

Although forgetting is often cited as a failure of memory, it is critical for efficient memory functioning (Nørby, 2015). Long-term memory represents a theoretically unlimited storage system from which memories must be accessed when necessary, relevant, or desired. This unlimited system poses a unique theoretical question about memory retrieval: If memory contains accumulated knowledge and experiences that may associate with other, related information, how can relevant information be accessed when desired, without being affected by interference from other memories? One potential solution is retrieval inhibition (Bjork, Bjork, & Anderson, 1998; Bjork, 1989), or activation of executive control processes that inhibit related, but currently irrelevant, information (Anderson, 2003, 2005). This can lead to successful remembering but can also have unintended consequences (Roediger, 1973; Tulving & Arbuckle, 1963). Specifically, attempting retrieval from long-term memory can sometimes cause forgetting of related information (Anderson, Bjork, & Bjork, 1994; Murayama, Miyatsu, Buchli, & Storm, 2014; see Storm & Levy, 2012, for a review). This phenomenon, termed *retrieval-induced forgetting*, is viewed as a negative consequence of memory retrieval (Bjork, 1975), in that information that is retrieved becomes subsequently more accessible, while related, but not retrieved, information becomes less accessible or temporarily forgotten.

Studies of retrieval-induced forgetting (RIF) typically employ a three-stage paradigm that includes a study phase, retrieval practice, and memory test (see Figure 1 for an example). During the study phase, participants learn a series of items drawn from different categories (e.g., clothing, appliances) presented as category-exemplar pairs (e.g., clothing – skirt, clothing – blouse, appliance – toaster). Once participants have studied these pairs, they move into a retrieval practice phase, during which a subset of studied exemplars are presented for guided retrieval. Participants are typically shown a category and item-specific cue (e.g., clothing – sk___) that serves to encourage retrieval of a specific studied item. After completing retrieval practice, participants' memory for all studied items is tested. This paradigm produces three types of test items: Practiced items from practiced categories (Rp+; e.g., skirt), unpracticed items from practiced categories (Rp-; e.g., blouse), and unpracticed items from unpracticed categories (Nrp; e.g., toaster). Generally, because of the benefits of retrieval practice, participants show better memory for practiced items from practiced categories (Rp+ items), relative to unpracticed items from unpracticed categories (Nrp items). However, participants also show worse memory for unpracticed items from practiced categories (Rp- items), relative to Nrp items, suggesting that retrieving some category exemplars induces temporary forgetting of related information (see MacLeod & Macrae, 2001). Importantly, this finding is not limited to *successful* retrieval practice, as even attempted retrieval can produce worse memory for Rp- items (Storm, Bjork, Bjork, & Nestojko, 2006; Storm & Nestojko, 2010).

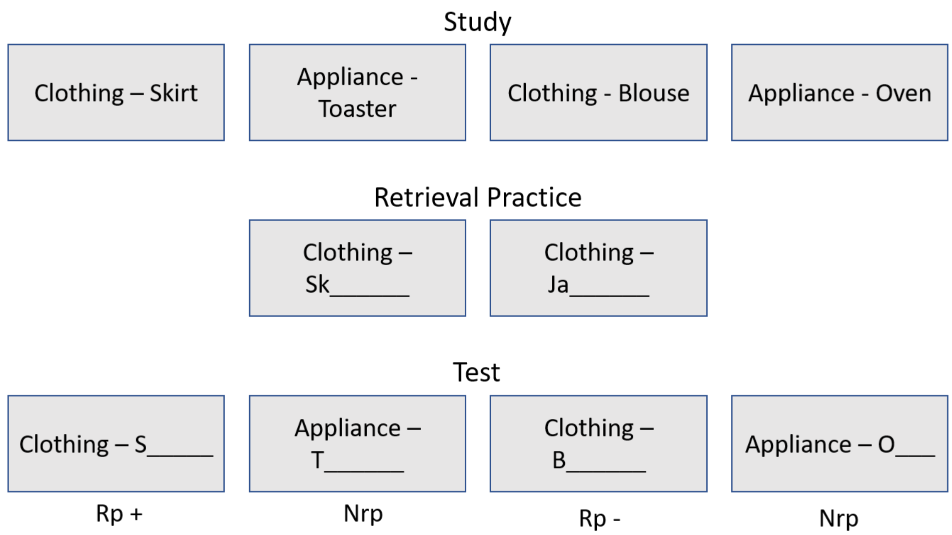

Figure 1. A sample schematic of a typical RIF paradigm. Retrieval practice produces Rp+ items (category exemplars that received practice), Rp- items (category exemplars that did not receive practice), and Nrp items (different category items that were not practiced).

RIF is a robust phenomenon, and has been documented using a wide variety of stimuli and scenarios, ranging from verbal materials (Anderson et al., 1994; Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007), autobiographical memories (Barnier, Hung, & Conway, 2004; Storm & Jobe, 2012), social cognition (Coman & Hirst, 2015; Macrae & MacLeod, 1999), visual scenes (Shaw, Bjork, & Handal, 1995), and factual statements (Anderson & Bell, 2001; Gómez-Ariza, Lechuga, Pelegrina, & Bajo, 2005). While the underlying mechanisms that support RIF are debated, inhibitory processes are often proposed as the most likely explanation (Anderson, 2003; Bjork, 1989; Storm & Levy, 2012). Inhibition serves a functional purpose to facilitate retrieval of relevant information, while simultaneously preventing interference from related, but currently irrelevant, information; the consequence of this process is that inhibited items become less accessible and therefore less recallable. In the context of an RIF paradigm, inhibition is engaged during retrieval practice to resolve competition between Rp+ and Rp– items because they are associated with the same retrieval cue (e.g., clothing). In this way, Rp+ items are successfully retrieved, and become more easily accessed during the memory test, while Rp- are inhibited and become less accessible and/or temporarily forgotten (Anderson, 2003, 2005; see also Storm, 2011). If retrieval is not engaged, inhibitory processes are unnecessary, an assumption that is supported by evidence that simply restudying a subset of material does not produce RIF (Anderson & Bell, 2001; Bäuml & Aslan, 2004; Ciranni & Shimamura, 1999; Staudigl, Hanslmayr, & Bäuml, 2010).

Although the majority of RIF research has used verbal study materials, and retrieval practice using tasks such as word stem completion, to investigate inhibition, recent work has shown that RIF also occurs for visual information (see also Ciranni & Shimamura, 1999; Fan & Turk-Browne, 2013; Waldhauser, Johansson & Hanslmayr, 2012; Shaw et al., 1995). For example, Maxcey and Woodman (2014) demonstrated that everyday objects are also subject to RIF. Participants studied a series of real-world objects grouped by category membership (e.g., muffins, cars). Because visual objects do not lend themselves to measures similar to word-stem completion, participants' retrieval practice consisted of a two-alternative-forced-choice (2AFC) task, during which they chose between one studied and one novel exemplar from a subset of studied categories. Across three experiments, Maxcey and Woodman (2014) observed robust RIF, with higher recognition performance for Rp+ exemplars and worse performance for Rp- exemplars, relative to Nrp objects. Importantly, this finding held even when participants completed a task that prevented object-specific naming strategies (Experiment 2), under articulatory suppression to prevent verbal coding (Experiment 3), and despite evidence that real-world objects are associated with extremely high recognition accuracy (Brady, Konkle, Alvarez, & Olivia, 2008; Standing, 1973). Research utilizing *recognition*-induced forgetting has expanded on this paradigm to show that semantic relatedness of objects, and not simply temporal grouping, is key to producing RIF for visual information (Maxcey, Glenn, & Stansberry, 2017). Combined, recognition- and retrieval-induced forgetting paradigms rest on the assumption that competition between category exemplars is critical to producing worse memory performance for Rp- items. In this way, many studies converge on the process of inhibition as an explanatory mechanism for RIF patterns observed.

Inhibition in RIF

There are several competing explanations for why RIF occurs. The inhibitory account of RIF (Anderson, 2003) suggests that when retrieval practice is engaged, competition between category members must be resolved to eliminate interference. To accomplish this, inhibitory processes act to suppress the representation of Rp- (unpracticed and currently irrelevant) items, resulting in poorer subsequent memory relative to Nrp items. Additionally, there are two non-inhibitory accounts of RIF. The associative blocking account assumes that retrieval practice serves to strengthen the bond between category labels and specific exemplars, and strengthened Rp+ items block access to Rp- items by producing interference during test (Raaijmakers & Jakab, 2013). Because Nrp items do not share the same category labels, these are unaffected. Third, context-based accounts hold that RIF arises due to mental context changes that occur during RIF paradigms (Jonker, Seli, & MacLeod, 2013). Specifically, participants engage retrieval practice under a different context than initial study, resulting in Rp+ items associated with two mental contexts. When retrieval practice context is reinstated during the final memory test, this improves memory for Rp+ items, but impairs memory for Rp- items because they do not have the same context association.

Although all three accounts have received empirical support (see Murayama et al., 2014, for a review), many studies converge on inhibitory processes as the most likely explanation for poor memory performance for Rp- items (but see Jonker et al., 2013; Raaijmakers & Jakab, 2013, for alternative explanations; these are considered in greater detail in the General Discussion). A majority of this evidence rests on the assumption that Rp- items are actively inhibited when retrieval of Rp+ items is attempted (Anderson, 2003; Anderson et al., 1994; Bjork, Bjork, & MacLeod, 2006; Levy & Anderson, 2002; for a review, see Storm & Levy, 2012). The inhibitory account of RIF assumes that when retrieval practice occurs, current category exemplars activate other, related items because they are associated with the same retrieval cue. This creates competition between items, which can be resolved by inhibiting the representation of Rp- items, therefore making them less accessible, and allowing the relevant representation to be selected. Inhibition is proposed to act either at the level of an item's semantic representation (e.g., Anderson, 2003; Anderson & Spellman, 1995; Johnson & Anderson, 2004), or episodic representation (Racsmány & Conway, 2006), with the ultimate goal to resolve interference.

There are four key predictions unique to the inhibitory account of RIF that have been supported by recent research. These predictions include cue independence, retrieval dependence, strength independence, and interference dependence. First, cue independence predicts that RIF should occur regardless of whether retrieval practices takes place utilizing identical cues (e.g., appliance – to____) or independent cues (e.g., bread – to____). For example, Anderson and Spellman (1995) demonstrated that retrieval practice using independent cues still produced RIF to the same extent as using identical cues. This prediction assumes that because inhibition acts at the level of an item's memory representation, retrieval of the item itself, and not retrieval of the cue, produces RIF (Anderson & Bell, 2001; Ciranni & Shimamura, 1999; Johnson & Anderson, 2004; Veling & Van Knippenberg, 2004; but see Perfect et al., 2004; Jonker et al., 2013).

Second, the inhibition account of RIF predicts retrieval dependence; this prediction is rooted in the assumption that RIF should only be observed when practice involves an *active* retrieval component. Specifically, simply re-exposing observers to a subset of studied items, or having observers perform their initial task (e.g., ratings) a second time should not produce RIF. This prediction is supported by research showing that a restudy component only produces facilitation for Rp+ items, but does not produce impairment for Rp– items (Ciranni & Shimamura, 1999; Experiment 5; Dobler & Bäuml, 2013). Similarly, retrieval practice wherein observers retrieve category names instead of category exemplars (e.g., app______ - toaster) does not produce RIF (Anderson, Bjork, & Bjork, 2000). Retrieval dependence suggests that the retrieval of *specific* items is necessary to activate inhibitory processes and therefore produce forgetting; without active retrieval, there is no need to resolve competition among items, and no RIF is produced (Anderson et al., 2000; Saunders, Fernandes, & Kosnes, 2009; but see Jakab & Raaijmakers, 2009).

Third, strength independence predicts that the magnitude of RIF is wholly dependent on how memorable the practiced (Rp+) items are, and not the degree to which Rp+ items are strengthened during retrieval practice. Evidence for strength independence was demonstrated by Macrae and MacLeod (1999), who observed that repeated retrieval practice does not affect the differences between Rp- and Nrp items. In addition, RIF is still observed even when retrieval is impossible (Storm et al., 2006). For example, Storm and Nestojko (2010) had observers learn category-exemplar pairs (e.g., appliance – toaster; fruit – lemon) before completing retrieval practice consisting of category names and two letter stem cues (e.g., appliance – to____). Importantly, in Experiment 3, a subset of practice items were inconsistent with any studied exemplars (e.g., appliance – di____ versus appliance – to____), and in Experiment 4, *all* practice items were inconsistent with studied exemplars (e.g., fruit – ma____ and fruit - ba____). Even under conditions in which retrieval practice was designed to be impossible, with most (Experiment 3) or all (Experiment 4) studied items from a particular category designated Rp-, the *attempt* to retrieve related category members still produced robust RIF for Rp- items, compared to Nrp items (see also Bäuml, 2002).

Finally, and critical to the current study, the inhibitory account of RIF proposes interference dependence. This prediction assumes that items that create larger degrees of competition (i.e., increased interference) during retrieval practice are more likely to require active inhibition of competing representations, therefore producing larger RIF. For example, Anderson et al. (1994) showed that items that were considered more semantically related (e.g., fruit – orange) were more susceptible to RIF, relative to weaker semantic relationships (e.g., fruit – tomato). Additionally, Storm, Bjork, and Bjork (2007) found that items that were episodically strong were most susceptible to RIF. Participants completed a directed-forgetting task, during which they were instructed that some studied items were to-be-remembered for the memory test, and others could be forgotten. Results showed that RIF for to-be-remembered items was considerably stronger than RIF for to-be-forgotten material.

Combined, these findings suggest that weak item representations do not produce significant competition during retrieval practice, and do not require the activation of inhibitory processes for retrieval success to occur; as a result, no RIF is observed (Anderson et al., 1994; see also Spitzer, 2014). However, strong representations that compete with Rp+ items are susceptible to forgetting because retrieval competition must be resolved to facilitate successful retrieval practice (Anderson et al., 2000; Bäuml & Samenieh, 2010, 2012; Shivde & Anderson, 2001; Sharman, 2011; Storm et al., 2007; but see Jakab & Raaijmakers, 2009).

Forgetting Visual Information

Although most RIF literature has examined memory for verbal material, recent work has expanded RIF and recognition-induced forgetting to visual material (e.g., Maxcey, 2016; Maxcey & Bostic, 2015; Maxcey, Bostic, & Maldonado, 2016). Recently, Reppa, Williams, Worth, Greville, and Saunders (2017) tested interference dependence using visual objects as stimuli. Visual stimuli lend themselves naturally to manipulations of item strength, given that everyday objects contain a wide range of both episodic detail (e.g., perceptual differences across within-category exemplars) and semantic detail (e.g., names, functions, likely locations, etc.). In particular, objects can be classified as stronger or weaker exemplars based on their prototypicality (Jolicoeur, Gluck, & Kosslyn, 1984; Rosch, Mervis, Gray, Johnson, & Boyes-

Braem, 1976; Tanaka & Taylor, 1991). Non-typical objects tend to be highly memorable because they do not share as many features with category members (Rosch et al., 1976; Tanaka & Taylor, 1991). However, typical objects, which share more features and are highly associated with category membership, should produce stronger competition (Anderson et al., 1994; Jonker et al., 2013). Reppa et al. (2017) had participants study typical and non-typical items before completing a recognition practice phase (Maxcey, 2016; Maxcey & Woodman, 2014). An important assumption of this recognition practice phase is that practiced exemplars produce spreading activation to other category members. However, this activation does not spread to all category-relevant items, only to those previously presented in experimental context (Maxcey & Woodman, 2014). In this way, spreading activation causes strong Rp- representations to compete with Rp+ representations; to resolve this, inhibitory processes are enacted to prevent irrelevant representations from being selected in favor of desired information.

Reppa et al. (2017) showed that object memorability influenced the magnitude of RIF: When participants practiced typical objects, memory for non-typical competitors was impaired, while practicing non-typical objects did not significantly impair memory for typical competitors. This finding, taken with other recent investigations of recognition-induced forgetting, suggests that recognition practice is sufficient to encourage retrieval competition (see also Maxcey, 2016; Maxcey et al., 2017). Second, and critical to the current study, these results suggest that baseline item strength directly influences RIF. In particular, these findings suggest that competition is less likely to be produced by larger numbers of shared features, or based on stronger category membership association (Anderson et al., 1994). Instead, retrieval competition may rely more on memory strength, with highly memorable items more susceptible to RIF when unpracticed, relative to weaker items.

Importantly, Reppa et al. (2017) argued that object typicality influences RIF based on the memorability of the item, and not strong association with category membership. For example, typical items are strongly associated with categories (e.g., apple with the category fruit) and are identified faster and more accurately than non-typical objects (e.g., dragonfruit or pomegranate; Jolicoeur et al., 1984; Murphy & Brownell, 1985; Rosch et al., 1976; Tanaka & Taylor, 1991). On the other hand, non-typical items tend to be highly distinct from other objects within the same category because of their unique visual features (e.g., a non-typical fruit, such as a dragonfruit, shares few features with an apple; Rosch et al., 1976; Tanaka & Taylor, 1991). Lack of shared physical characteristics increases non-typical objects' memorability. This in turn makes non-typical objects stronger competitors when they are not retrieved and necessitates the activation of inhibitory processes.

Beyond the novel support for interference dependence, visual objects in a RIF paradigm allow for assessment of how representations might change as a consequence of inhibitory processes. If, as the inhibitory account proposes, representations of Rp- items are suppressed when they compete with Rp+ items, this should result in weakening of their memory representations. One assumption of object typicality is that unique features separate items from each other in psychological space (Johnston, Milne, Williams, & Hosie, 1997; Lee, Byatt, & Rhodes, 2000; Valentine, 1991), which leads to higher discriminability and better memory. If inhibitory processes serve to weaken representations in memory, this may also be reflected by changing representations in psychological space.

Measuring Psychological Space

Representations in psychological space can be quantified using measures that assess perceived similarity between items, such as multidimensional scaling (MDS; Kruskal, 1964;

Shephard, 1962; see Hout, Papesh, & Goldinger, 2013, for a review). MDS describes a set of statistical procedures that can be used for dimension reduction; specifically, MDS takes input, such as similarity ratings, and reduces the number of variables under consideration to allow for a set of principal dimensions (e.g., the most useful feature dimensions used to discriminate between category exemplars) to be assessed. By applying these procedures to estimates of similarity, the distances between items can be quantified, with greater similarity indicated by less distance. The application of MDS to object perception also allows for a visual appreciation of the underlying relations between groups of objects, and has been used previously to demonstrate differences between more and less distinct visual material (e.g., faces, Faerber, Kaufmann, Leder, Martin, & Schweinberger, 2016; Papesh & Goldinger, 2010; Valentine, 1991; everyday objects, Hout, Goldinger, & Brady, 2014). Across these studies, objects that are atypical or unusual to observers are less clustered in psychological space and have lower similarity ratings, while those that are more typical show greater clustering and higher similarity.

To assess similarity, researchers can employ either direct or indirect measures to elicit similarity estimates about stimuli (Jaworska & Chupetlovska-Anastova, 2009). Direct measures include overt decisions, such as classifying items into categories (e.g., Faye et al., 2004; 2006), with similarity data measured by how often stimuli are classified as the same, or different, category. Indirect measures often involve secondary measurements, such as same/different judgments about two exemplars. Using these, similarity can be estimated based on the percentage of items that are mistakenly judged *same* (e.g., Wish & Carroll, 1974), or by response speed, with slower responses indicating higher similarity (e.g., Papesh & Goldinger, 2010). Most often, similarity is assessed by having observers rate all stimuli in pairs, with a rating for every possible combination of stimuli. However, this is not always the optimal method for data collection, as the number of ratings participants must make increases dramatically with the number of stimuli in the set. To accommodate this, Hout, Goldinger, and Ferguson (2013) developed an MDS method that relies on the spatial relations between items (Spatial Arrangement Method, SpAM; see also Goldstone, 1994). Using this method, participants move stimuli around the computer screen, with distance between items serving as a measure of similarity. Critically, this method has been shown to produce MDS solutions that are comparable to those generated by traditional pairwise comparisons, making it both effective and efficient to collect similarity data for visual materials.

An important assumption of similarity metrics such as MDS, and the accompanying psychological spaces, is that similar items are closer together. Limited space between representations makes them more confusable during retrieval, contributing to less accurate recognition. This concept is related to inhibitory mechanisms proposed to underlie RIF: If items are closer together in space, they will also produce greater competition during retrieval. However, this is inconsistent with evidence that highly memorable material is more susceptible to forgetting than non-memorable (i.e., typical) material (Reppa et al., 2017), given that memorability should distance items from each other (Johnston et al., 1997; Lee et al., 2000; Valentine, 1991). Therefore, assessing RIF using visual stimuli allows for a unique assessment of how memory strength (e.g., non-typical exemplars associated with stronger memory traces) and category associations (e.g., typical exemplars are more strongly associated with category membership) contribute to RIF, although to date, few studies have empirically investigated RIF using visual material as stimuli.

Current Study

The aim of the current study was to further investigate the inhibitory explanation for RIF, specifically by examining interference dependence using visual stimuli that vary naturally in memory strength. The current study converges on several broad hypotheses about how memory strength affects susceptibility to RIF. First, assumptions of cue independence and strength independence posit that inhibitory mechanisms act at the level of an item's representation (Anderson, 2003), although there is mixed evidence about whether inhibition acts on semantic information (Anderson, 2003; Anderson & Spellman, 1995; Johnson & Anderson, 2004) or episodic information (Racsmány & Conway, 2006). However, the ultimate goal of these inhibitory processes is to resolve interference. Second, assumptions of interference dependence hold that items that cause the greatest degrees of competition during retrieval practice will show the most subsequent forgetting.

Objects make an ideal class of stimuli with which to test assumptions about interference dependence because typical and non-typical objects differ in their episodic strength, but may contain comparable semantic information. There is growing evidence that baseline item strength predicts the presence and magnitude of RIF (Reppa et al., 2017; Spitzer, 2014), with stronger category exemplars producing greater RIF when unpracticed competitiors. If non-typical objects associated with stronger memory traces are more susceptible to forgetting when they are unpracticed competitors, this would provide additional evidence of interference dependence that cannot be easily explained by other theories of RIF (e.g., context-based accounts; Jonker et al., 2013). If non-typical objects are more susceptible to forgetting, this would provide evidence that stronger memories compete more during practice, and are more likely to be subject to RIF.

Finally, assessing memory through alternative measures, such as by measuring perceived similarity in psychological space, provide a novel means to test the inhibitory perspective of RIF. Core to this argument is that Rp- items are inhibited, or their activation suppressed, to resolve response competition during retrieval practice. If Rp+ items receive enhanced representation due to the benefits of retrieval, while Rp- items are inhibited and receive diminished representation, this should be reflected by increased and decreased distances in psychological space relative to Nrp items, repsectively. MDS also allows for a critical test of the hypothesis that episodic representations contribute to RIF (Racsmány & Conway, 2006): If perceptual similarity creates more competition between category exemplars, as would be the case for typical objects, the magnitude of RIF should increase for typical items. Alternatively, if memory strength creates addtional competition between category exemplars, as would be the case for non-typical objects, the magnitude of RIF should increase for non-typical items. See Table 1 for a breakdown of the hypotheses and predicted results for Experiment 1, and Table 2 for a similar breakdown for Experiment 2.

Table 1. Hypotheses and Predicted Results as applied to Experiment 1.

Hypothesis	**Predicted Results**
1. Stronger memories produce additional competition when unpracticed.	Non-typical Rp- objects, which are highly memorable, should show RIF.
2. Perceptually similar items produce additional competition when unpracticed.	Typical Rp- objects, which are perceptually similar, should show RIF.
3. Perceived similarity in psychological space will change as a function of RIF.	Rp- objects will show smaller distances (i.e., appear more similar) relative to Nrp items.

Experiment 1. Forgetting Objects

The goal of Experiment 1 was to replicate and extend the recent finding that visual materials are susceptible to recognition-induced forgetting (Reppa et al., 2017), by comparing RIF across typical and non-typical objects. Experiment 1 consisted of a three-stage task, during which participants were exposed to non-typical and typical objects grouped by categories (e.g., backpacks, lamps), practiced or re-studied a subset objects, including non-typical and typical objects from different categories, and finally, the contents of memory assessed. To extend previous work, participants in Experiment 1 experienced one of two possible final memory tests. In Experiment 1a, participants performed a standard recognition memory test to examine whether typical and non-typical objects show RIF based on category membership. In Experiment 1b, participants' memory was assessed by multidimensional scaling (MDS).

Across Experiments 1a and 1b, there were several key predictions related to the inhibitory account of RIF. First, a core assumption of the inhibitory account of RIF holds that when subsets of semantically related items are practiced, representations of unpracticed, but related, items are inhibited (Anderson, 2003). This inhibition results in sharpening of Rp+ representations, while weakening Rp- representations. Based on evidence that semantically related visual materials are also subject to RIF (e.g., Maxcey & Woodman, 2014; Maxcey et al., 2017; Reppa et al., 2017; Rugo et al., 2017), Experiment 1 examined RIF across typical and non-typical objects grouped by category membership, with the prediction that participants would show RIF for Rp- exemplars.

A second assumption of the inhibitory account of RIF is interference dependence, which holds that items that create larger degrees of competition produce larger magnitude RIF (see Anderson, 2003; Storm & Levy, 2012). Previous work has shown that words that are high frequency and have a strong semantic association with category cues (Anderson et al., 1994; Shivde & Anderson, 2001), and items that are episodically strong (Storm et al., 2007), produce larger differences between Rp+ and Rp- exemplars. These findings led to the prediction that non-typical objects, which share fewer features with other category exemplars, should show *less* RIF when not practiced, relative to typical items. However, this is inconsistent with recent evidence that items that are highly memorable, such as non-typical objects, may be more susceptible to RIF when not practiced (e.g., Reppa et al., 2017; Rugo et al., 2017; see also Spitzer, 2014). Instead, recent evidence suggests that baseline item strength, and not category association, may be a better predictor of RIF (Spitzer, 2014). This would suggest that non-typical items, which are highly memorable to observers based on their unique characteristics, would produce larger degrees of competition during retrieval practice and ultimately a larger magnitude RIF when they are non-practiced competitors.

To test the assumption that interference dependence relies more heavily on baseline memory strength than shared features, Experiment 1 employed typical objects (i.e., those that are more strongly associated with category membership and contain multiple shared features, but are less memorable) and non-typical objects (i.e., those that may be weakly associated with category membership and contain fewer shared features, but are highly memorable) as studied and practiced items. Importantly, participants only practiced one type of object (typical or non-typical) from studied categories. If highly memorable Rp- items produce additional competition during retrieval of Rp+ items, this should produce larger degrees of forgetting for non-typical items. This prediction is consistent with recent evidence that strong memorial representations produce more competition when weaker representations are retrieved, leading to larger RIF for memorable Rp- items (Reppa et al., 2017). Alternatively, if shared features contribute more to competition during retrieval practice, with strong associations between category and exemplar activating Rp- representations, the degree of observed forgetting should be larger for non-practiced *typical* items.

Lastly, inhibitory processes in RIF arguably serve to resolve competition between memory representations in order to facilitate retrieval (Storm & Levy, 2012). One way to quantify the degree of sharpening or inhibition among memory representations is by employing alternative measures of memory processes, including MDS. By employing MDS in a RIF paradigm, visual representation of relational structures between practiced and unpracticed category members can be examined. This led to a third hypothesis, specific to Experiment 1b: If RIF improves memory for Rp+ items through the inhibition of Rp- items, Rp- items should appear more similar to each other in psychological space. This hypothesis would be supported by smaller average inter-item distances for Rp- items, relative to baseline items.

Experiment 1a: Method

Participants

Based on an a priori matched pairs analysis of Reppa et al. (2017), who demonstrated recognition-induced forgetting of typical and non-typical objects, approximately 60 participants were necessary for Experiment 1, divided between Retrieval Practice and Control instructions. This estimate was based on mean group differences across non-typical and typical objects, with significant RIF observed for non-typical objects, compared to the nonsignificant RIF for typical objects, with Cohen's d = 0.64, α = 0.05, and power held at 0.95. In total, 78 participants from Louisiana State University completed the experiment in exchange for partial course credit. Forty-five participants (M_{age} = 19.4 years, 37 female) were randomly assigned to Retrieval Practice conditions, and the remaining 33 (M_{age} = 19.8 years, 26 female) were assigned to a Control

(Restudy) condition. All participants self-reported normal or corrected-to-normal vision, normal color vision, and were native English speakers. Participants engaged in individual sessions lasting no more than 45 minutes.

Stimuli

There were 365 objects available for the current experiment. Each participant viewed a subset of 96 everyday objects, drawn evenly from 12 distinct categories. Each category contained a maximum of 17 possible exemplars, and all categories were drawn from the Massive Memory MDS Database (Hout et al., 2014). Images appeared in greyscale to encourage participants to consider typicality on the basis of shape only. Participants studied typical and non-typical exemplars from each category. There were 10 possible additional categories that served as filler; these were not tested or analyzed. To identify a suitable set of typical and non-typical objects, images were identified from MDS similarity ratings made by naïve observers (Hout et al., 2014). Images consistently rated as atypical or unusual, as indexed by low typicality rankings, and located toward the fringes of psychological space, were used as non-typical objects. Images consistently rated as prototypical, as indexed by high typicality rankings, and located toward the center of psychological space, were used as typical objects. Sample stimuli are illustrated in Figure 2, and available categories are listed in Appendix A. During the first phase of the experiment, participants provided pleasantness ratings for the two most and least typical exemplars from studied categories. During retrieval practice, participants were presented with novel category- and typicality-consistent distractors (e.g., if typical birdhouses were practiced, two novel, typical birdhouses served as distractors). During the final memory test, participants were again presented with novel category- and typicality-consistent distractors that were not viewed during any previous phase of the experiment, intermixed with filler objects.

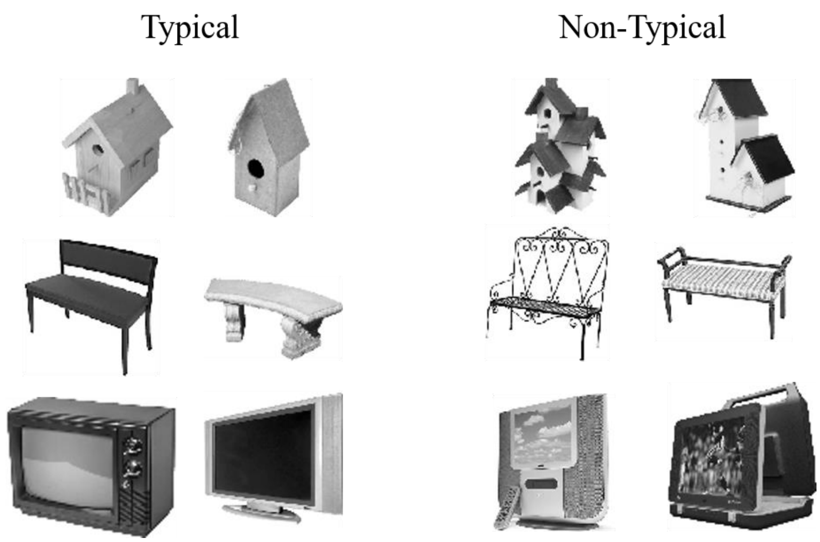

Figure 2. Sample images used in Experiment 1a and 1b. Left are typical objects, and right are non-typical objects. Categories include birdhouses, benches, and televisions.

Design

Participants were divided based on Recognition Practice (henceforth, Practice group) or Control instructions (i.e., restudy with no active retrieval component). The design for the Practice group was repeated measures, with participants experiencing two levels of Typicality (typical, non-typical), and tested on three Item Types (Rp+, Rp-, Nrp). Rp+ items were objects that were practiced during the recognition practice phase, Rp- items were studied objects that were unpracticed, but related to practiced categories, and Nrp items were studied objects that were unrelated to practiced categories and also unpracticed. For the Control group, participants experienced no recognition practice and were simply re-exposed to a subset of study items, which should not produce RIF (Anderson et al., 2000; Ciranni & Shimamura, 1999, Experiment 5; Dobler & Bäuml, 2013). The design for the Control group was also repeated measures, with participants experiencing two levels of Typicality, and tested on three Item Types (S+, S-, Ns). S+ items were objects that were studied and restudied, S- items were objects that were studied and not restudied, but related to restudied categories, and Ns items were objects that were not restudied and unrelated to restudied categories. Practiced categories and Typicality (typical, non-typical) were counterbalanced across subjects, with participants engaging in retrieval practice or restudy with different typical and non-typical objects.

The dependent variables for Practice group participants included hit rates for practiced (Rp+), unpracticed but related (Rp-), and baseline (Nrp) items, as well as participants' false alarms to novel category- and typicality-consistent distractors. The dependent variables for Control group participants included hit rates for restudied (S+), not restudied but related (S-), and non-restudied (Ns) items, and false alarms. In addition, all participants' accuracy as *A′* scores was assessed as a measure of discriminability, due to its resistance to very high or low scores (see Rugo et al., 2017).

Procedure

Practice group participants experienced three experimental phases, including study, recognition practice, and test; Control group participants also experienced three experimental phases, including study, restudy, and test. Each phase is described in detail below. See Figure 3 for a schematic of the study and recognition practice phase, and Figure 4 for a schematic of the test phase.

Study Phase. Because observers have been shown to have highly accurate and detailed memory for objects (Brady et al., 2008), participants engaged in an incidental learning task during the study phase (see Reppa et al., 2017). During the study phase, Practice and Control participants began each trial with a 500 ms fixation cross. They then viewed individual objects and indicated how pleasant they found the object on a 1 to 5 scale, with “1” indicating that the object was very unpleasant, and “5” indicating that the object was very pleasant (Reppa et al., 2017). Participants studied 16 typical and 16 non-typical objects, 4 from each category. To minimize the influence of primacy and recency effects (Murdock, 1962), four additional items from novel categories were presented at the beginning and end of the study phase, for a total of 40 rated objects. Each object was followed by a 500 ms inter-trial interval. Upon completion of the study phase, participants experienced a 5-minute break, during which they performed a simple visual search task. This task involved searching for Waldo in scenes digitally adapted from *Where's Waldo?* books (Maxcey, 2016; Rugo et al., 2017).

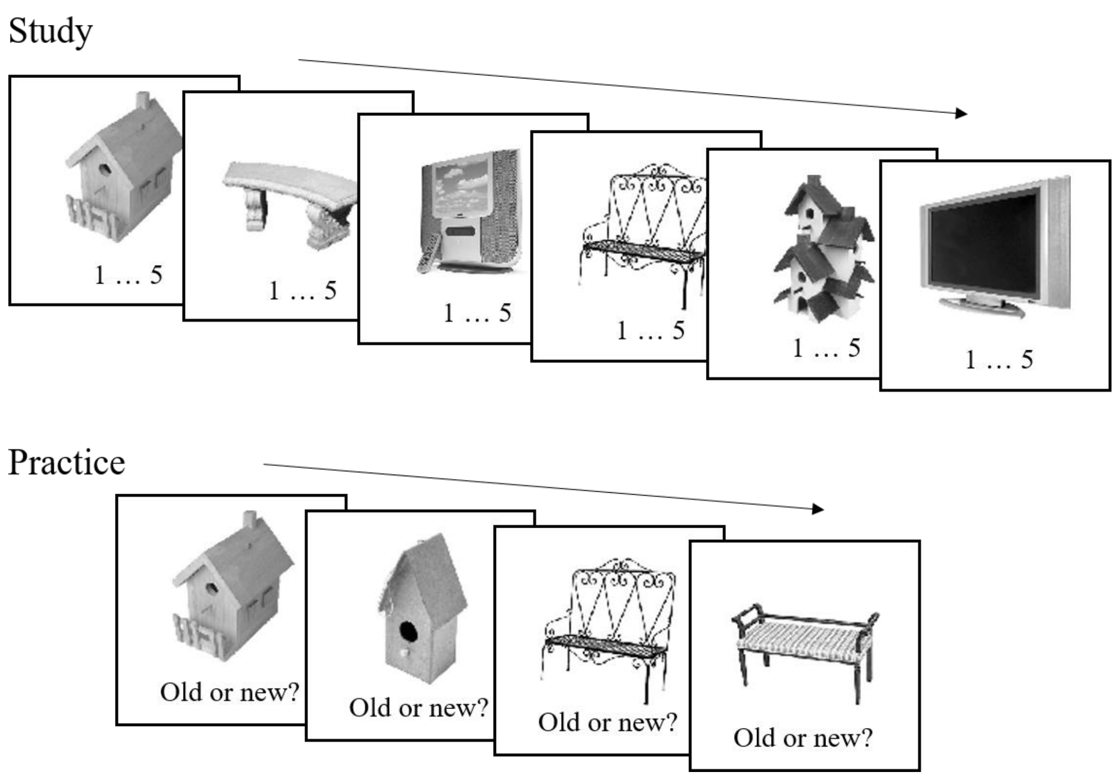

Figure 3. A schematic of the study phase, during which participants made pleasantness ratings about everyday objects, and the retrieval practice phase, during which participants completed old/new recognition decisions. During study, participants rated typical and non-typical objects from each category. During retrieval practice, participants practiced only typical or only non-typical objects from half of the studied categories.

Recognition Practice Phase. Practice group participants experienced a second phase during which they performed a recognition memory test on half of the studied objects from half of the studied categories. Although participants studied typical and non-typical objects from all categories, typical and non-typical practiced objects were drawn from different categories. On each trial, a single object was presented, and participants indicated whether the object was "old" (i.e., previously studied) or "new" (i.e., novel to the experiment). Participants completed a total of 24 recognition practice trials, with eight previously studied items randomly intermixed with eight category- and typicality-consistent distractor objects, and eight filler objects from novel categories (see, e.g., Figure 3). Within each studied category, participants practiced either typical or non-typical exemplars (e.g., two typical backpacks, two non-typical vases, etc.). Each trial was followed by feedback for 500 ms indicating whether the answer was correct or incorrect.

Although other previous work has used two-alternative forced choice (2AFC) during retrieval practice (e.g., Maxcey & Woodman, 2014; Maxcey, 2016), participants in the current study made single-item recognition decisions (Reppa et al., 2017). This practice task was intended to induce competition between category exemplars by encouraging recollection-based recognition (e.g., Brown, 1976; Mandler, 1980). Specifically, by including distractor items that were similar to studied items (e.g., two typical birdhouses) and by using a yes/no recognition task, participants must make practice responses on the basis of recollection, rather than relying on familiarity (Bodner & Richardson-Champion, 2007). In the case of 2AFC decisions, detailed memory may not be necessary for successful retrieval, as observers may rely on familiarity between two choices or gist-based representations (see Cunningham, Yassa, & Egeth, 2015). Additionally, the use of corrective feedback during the practice phase was intended to ensure that practice of target items was effective (Raaijmakers & Jakab, 2013; Reppa et al., 2017).

Additionally, the use of single-item recognition practice encourages competition based on the assumption that to make a decision about an item’s memory status, a search-like competitive memory process must be engaged (Brown, 1976; Mandler, 1980; see Reppa et al., 2017). This search process involves practice items' category associations; although participants are not explicitly taught to associate exemplars with cues, as is the case in investigations of RIF using verbal stimuli, they nevertheless experience implicit cue associations through groupings of semantically related items. The competition produced is therefore based in sharing of an implicit category cue, allowing for a direct test of the hypothesis that category association may be predictive of RIF for visual information.

Participants completed three recognition practice phases, with objects shown in a different, randomly assigned order for each iteration (Reppa et al., 2017). After each practice block, and upon completion of the second phase, participants again performed a visual distractor task by searching *Where's Waldo* scenes. Participants in the Control condition simply restudied three times, in random order, the same number of practiced objects intermixed with eight novel and category-consistent distractors and eight filler objects.

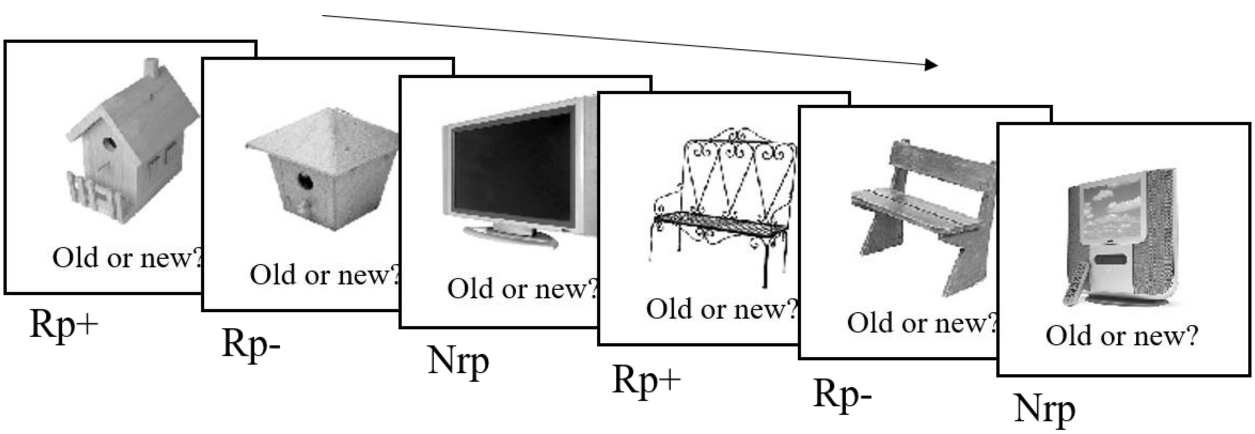

Figure 4. Trial schematic of the final recognition test. Participants made old/new recognition decisions about studied and practiced items (Rp+), studied items related to practiced categories (Rp-), and studied but unpracticed items (Nrp). The first three images represent typical objects; the final three represent non-typical objects.

Test Phase. All participants completed a test phase, during which they made old/new recognition judgments about studied and unstudied objects. All 32 objects from the study phase were intermixed with the same number of novel objects, four drawn from each category but that had not been presented during any previous phases, and with 12 filler objects, for a total of 76 recognition decisions. Presentation order of objects was fully randomized. Practice group participants experienced three types of studied objects: Rp+ objects (practiced), Rp- objects (unpracticed, related), and Nrp objects (unpracticed, unrelated). Participants in the Control group also experienced three types of studied objects that corresponded with retrieval practice items: S+ (restudied), S- (not restudied, related), and Ns objects (not restudied, unrelated). Participants indicated whether an object was previously rated during the first phase of the experiment (i.e., old), or if the object was novel (i.e., new).

Experiment 1a: Results

Alpha for significance tests was held at .05, and all multiple comparisons were Bonferroni-corrected unless stated otherwise. Only data from the eight experimental object categories were analyzed for each participant.

Recognition Practice

To ensure that participants in the Practice group successfully completed the recognition practice phase, hits (defined as correct recognition of studied objects), false alarms, sensitivity as *A′* scores and bias as *B″* were assessed by separate paired samples *t*-test across typical and non-typical practiced objects.

Participants showed slightly numerically higher hit rates for non-typical studied exemplars (*M* = .84, *SE* = .03) than typical studied exemplars (*M* = .82, *SE* = .03), but the difference was not statistically reliable, *t*(44) = .82, *p* > .05, Cohen's *d* = .12. However, participants did exhibit a reliably lower false alarm rate for non-typical distractor objects (*M* = .32, *SE* = .03), relative to typical distractor objects (*M* = .50, *SE* = .03), *t*(44) = 4.88, *p* < .05, Cohen's *d* = .73. Further, participants' *A'* scores were also reliably different, with participants exhibiting greater sensitivity for non-typical objects (*M* = .83, *SE* = .02), compared to typical objects (*M* = .73, *SE* = .03), $t(44) = 3.14$, $p < .05$, Cohen's $d = .47$; see Figure 5 for an illustration of false alarm rates and *A'* scores. There was no reliable difference in *B''* scores across non-typical (*M* = -0.44, *SE* = .07) and typical (*M* = -0.61, *SE* = .06) exemplars, $t(44) = 1.82$, $p > .05$, Cohen's $d = .27$. Finally, I expected to find that hits during the recognition practice phase would be significantly above chance for both types of practiced objects (Rugo et al., 2017). Hit rates for non-typical and typical objects were both significantly above chance, $t(44) = 18.23$, $p < .05$, Cohen's $d = 2.72$, and $t(44) = 12.70$, $p < .05$, Cohen's $d = 1.89$, respectively. Although participants did not show reliably higher hit rates for non-typical items, false alarm rates and *A'* scores support the hypothesis that non-typical objects are more memorable than typical objects (Reppa et al., 2017).

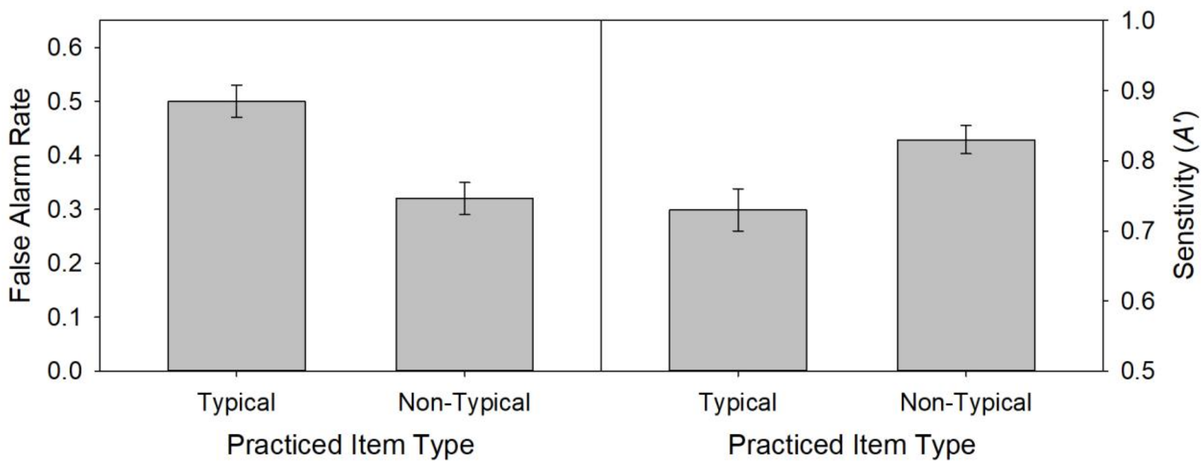

Figure 5. False alarm rates (left) and sensitivity (*A'*) across typical and non-typical items for participants who experienced retrieval practice. Error bars represent standard error of the mean (SEM).

Recognition Test

To assess participants' memory performance, hits (again defined as correct recognition of studied objects), false alarms to novel category- and typicality-consistent distractors, sensitivity as *A'* scores, and bias as *B''* scores was examined. Summary statistics across hits, sensitivity, and bias are available in Table 2 for participants who received Practice instructions, and Table 3 for participants who received Control instructions.

Table 2. Hits, sensitivity as A' scores, and bias as B" scores for participants who experienced Practice conditions.

	Nrp		Rp-		Rp+	
	Typical	**Non-Typical**	**Typical**	**Non-Typical**	**Typical**	**Non-Typical**
Hits	.67 (.03)	.70 (.03)	.58 (.03)	.48 (.03)	.79 (.03)	.92 (.03)
A'	.70 (.03)	.84 (.02)	.64 (.04)	.61 (.05)	.79 (.02)	.89 (.02)
B"	-.07 (.07)	.19 (.08)	.08 (.08)	.52 (.07)	-.37 (.09)	-.60 (.09)

First, to assess whether all participants demonstrated RIF regardless of the study material, hits (defined as correct recognition of studied objects) for participants in the Practice group were assessed by a 3 (Item Type: Rp +, Rp -, Nrp) x 2 (Typicality: Typical, non-typical) factor, repeated measures (RM) ANOVA. I predicted that, regardless of the studied category's Typicality, participants should show significant RIF, with better memory for studied objects (Rp+), relative to baseline objects (Nrp), and worse memory for related but unstudied objects (Rp-), relative to baseline objects (Nrp). This prediction was consistent with previous evidence that objects are not immune to recognition-induced forgetting (Maxcey & Woodman, 2014; Reppa et al., 2017). Participants' hit rates showed a significant RIF effect, with reliable differences based on Item Type, $f(2, 86) = 55.4$, $p < .05$, $\eta^2 = .56$. Planned comparisons revealed that hit rates for Rp- items ($M = .53$, $SE = .03$) were reliably lower than Nrp items ($M = .68$, $SE = .03$), $p < .05$, Cohen's $d = .71$, and Rp- and Nrp hit rates were both reliably lower than for Rp+ items ($M = .85$, $SE = .03$), *p*s < .05, Cohen's $d = 1.56$ and $d = .85$, respectively. Although there was no reliable difference across Typicality, $F(1, 43) = .45$, $p > .05$, $\eta^2 = .01$, the main effect of Item Type was qualified by a significant interaction, $F(2, 86) = 6.71$, $p < .05$, $\eta^2 = .56$. This interaction, illustrated in Figure 6, demonstrated that the RIF effect in hit rates was larger for non-typical exemplars than it was for typical exemplars, supporting the hypothesis that more memorable information is more susceptible to RIF (Reppa et al., 2017).

To examine whether RIF in hit rates was observed only for Non-Typical objects, separate

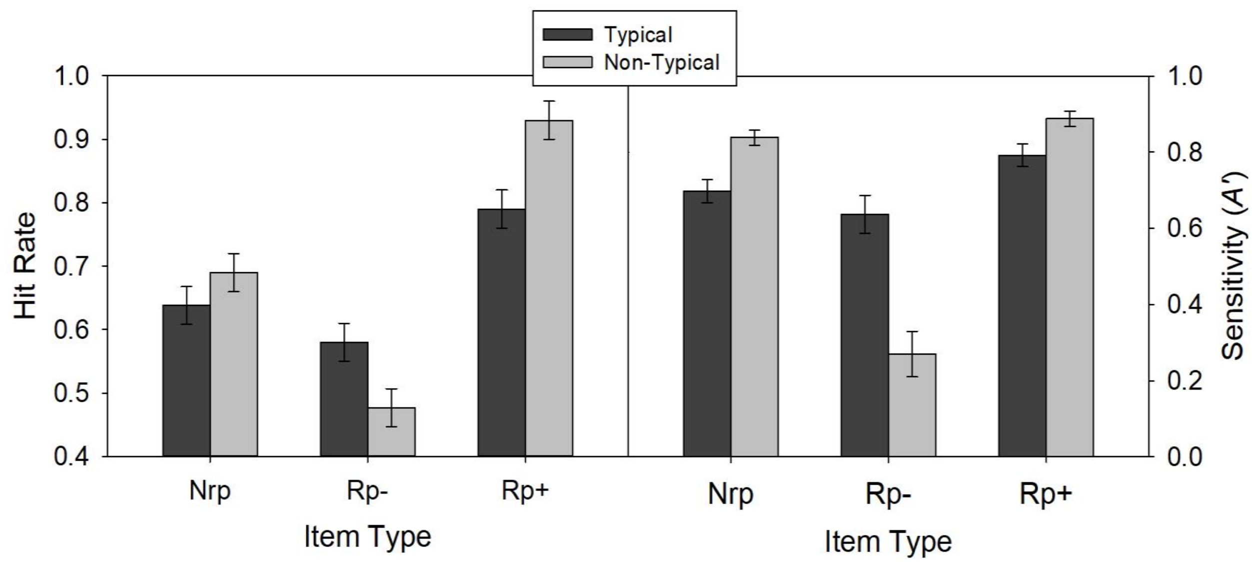

Figure 6. Hit rates (left) and sensitivity (*A'*, right) across Nrp, Rp-, and Rp+ items as a function of whether the exemplar was typical or non-typical. Error bars represent SEM.

3-factor RM ANVOAs across Item type as a function of Typicality were conducted. These analyses yielded reliable differences across both Typical exemplars, $F(2, 86) = 10.74$, $p < .05$, $\eta^2 = .20$, and Non-typical exemplars, $F(2, 86) = 59.72$, $p < .05$, $\eta^2 = .58$. However, planned comparisons revealed only Non-typical items showed a reliable RIF effect: Hit rates for Non-typical Rp+ items were reliably higher than hits for Nrp items, and Nrp items were reliably higher than Rp- items, *p*s < .05. However, hit rates for Typical items only showed reliable differences between Rp+ items, relative to both Nrp and R- items, $p < .05$, while Nrp and Rp-items did not reliably differ, $p > .05$. This directly replicates Reppa et al.'s (2017) finding that item strength modulates RIF effects.

Hits for participants in the Control group were also assessed by a 3 (Item Type: S+, S-, Ns) x 2 (Typicality) RM ANOVA. Unlike the Practice group, I predicted that there should be no RIF; instead, Control group participants should show only facilitation for restudied items (S+),

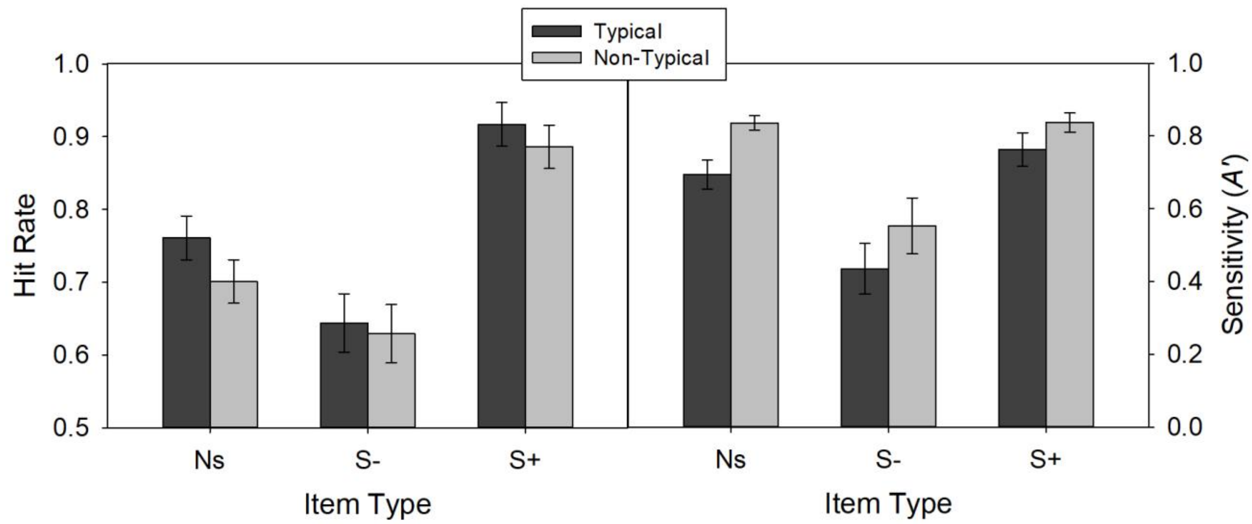

Figure 7. Hit rates (left) and sensitivity (*A′*, right) across Ns, S-, and S+ items as a function of whether the exemplar was typical or non-typical. Error bars represent SEM.

and no differences between unstudied but related items (S-) and baseline items (Ns). Similar to participants in the Practice group, Control participants' hit rates showed a main effect of Item Type, $F(2, 64) = 34.83$, $p < .0$, $\eta^2 = .52$, and no effect of Typicality, $F(1, 32) = 1.62$, $p > .05$, $\eta^2 = .05$. Unlike the Practice group, the interaction was not reliable, $F(2, 64) = .26$, $p > .05$, $\eta^2 = .01$. Planned comparisons revealed that Control group participants' hit rates were reliably different across all item types, with participants showing lowest hits for S- items ($M = .64$, $SE = .03$, $p < .05$) compared to Ns items ($M = .73$, $SE = .03$), $p < .05$, Cohen's $d = .52$, and higher hits for S+ items ($M = .90$, $SE = .03$) compared to Ns items $p < .05$, Cohen's $d = .97$. This pattern, illustrated in Figure 7, suggests restudy may produce similar RIF effects as retrieval practice for visual stimuli, and does not entirely support the prediction that restudy fails to produce RIF (Anderson & Bell, 2001; Bäuml & Aslan, 2004; Ciranni & Shimamura, 1999; Staudigl et al., 2010).

To investigate whether the magnitude of the observed RIF effect was larger for participants who received Practice instructions than Control instructions, hit rates were also assessed by a 3 (Item Type) x 2 (Typicality) x 2 (Condition: Practice, control) mixed model ANOVA, with Condition held between subjects. This analysis revealed a significant interaction between Item Type, Typicality, and Condition, $F(2, 150) = 3.76$, $p < .05$, $\eta^2 = .05$. Importantly, post-hoc tests revealed that Practice group participants showed significantly lower hit rates for Non-typical Rp- items ($M = .48$, $SE = .03$), relative to Control participants' corresponding Non-typical S- items ($M = .63$, $SE = .05$), $t(75) = 2.73$, $p < .05$, Cohen's $d = .63$. The difference between Typical Rp- items ($M = .58$, $SE = .03$) and Typical S- items ($M = .64$, $SE = .05$) was numerical only, $t(75) = 1.13$, $p > .05$, Cohen's $d = .26$. These findings provide additional support to the conclusion that RIF effects are produced by retrieval practice, and that they are larger for memorable (i.e., non-typical) information.

Second, false alarm rates were assessed by separate paired samples *t*-tests across both Practice and Control group participants, to examine whether participants showed higher false alarm rates to typical, relative to non-typical, items. Practice group participants showed reliably higher false alarm rates for typical distractors ($M = .36$, $SE = .02$), relative to non-typical distractors ($M = .21$, $SE = .03$), $t(43) = 7.43$, $p < .05$, Cohen's $d = 1.11$. Similarly, Control group participants showed higher false alarm rates for typical distractors ($M = .52$, $SE = .03$), relative to non-typical distractors ($M = .29$, $SE = .02$), $t(32) = 8.05$, $p < .05$, Cohen's $d = 1.40$; see Figure 8 for a comparison across Practice and Control instructions. Considered with participants' hit rates, these findings support the hypothesis that non-typical objects are more memorable, and may produce more competition during retrieval practice.

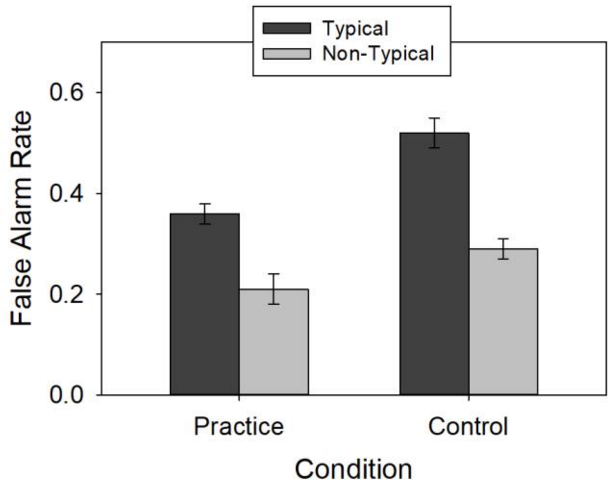

Figure 8. False alarm rates across Practice (left) and Control (right) conditions, as a function of test items' typicality. Error bars represent SEM.

Next, *A'* scores were examined by a 3 (Item Type: Rp+, Rp-, and Nrp) x 2 (Typicality: Typical, non-typical) RM ANOVA for Practice group participants. For Practice group participants, this analysis yielded a significant effect of Item Type, $F(2, 86) = 5.42$, $p < .05$, $\eta^2 = .11$, a reliable effect of Typicality, $F(1, 43) = 5.68$, $p < .05$, $\eta^2 = .12$, and a reliable interaction, $F(2, 86) = 4.22$, $p <.05$, $\eta^2 = .09$. Planned comparisons demonstrated a similar pattern as in participants' hit rates: *A'* scores for Rp- items were reliably lower (*M* = .45, *SE* = .05), followed by scores for Nrp items (*M* = .77, *SE* = .03), $p < .05$, Cohen's *d* = .52. *A'* scores for Rp+ items were reliably higher than both Rp- and Nrp items (*M* = .84, *SE* = .02), *p*s < .05, Cohen's *d* = .41 and .44, respectively. The significant interaction between Item Type and Typicality are illustrated in the right panel of Figure 6, and demonstrated a larger magnitude RIF for non-typical objects (Reppa et al., 2017). Like hit rates, separate 3-factor RM ANVOAs across Item type as a function of Typicality yielded reliable differences across both Typical exemplars, *F*(2,

86) = 5.56, $p < .05$, $\eta^2 = .12$, and Non-typical exemplars, $F(2, 86) = 19.76$, $p < .05$, $\eta^2 = .32$. Once again, only Non-typical items showed a reliable RIF effect in *A'* scores: *A'* for Non-typical Rp+ items were reliably higher than hits for Nrp items, $p = .05$, Cohen's $d = .37$, and Nrp items were reliably higher than Rp- items, $p < .05$, Cohen's $d = .57$. However, *A'* for Typical items only showed reliable differences between Rp+ items, relative to both Nrp and Rp- items, $p < .05$, Cohen's $d = .31$ and .65, respectively, while Nrp and Rp- items did not reliably differ, $p > .05$, Cohen's $d = .16$.

Table 3. Hits, sensitivity as *A′* scores and bias as *B″* scores for participants who received Control instructions.

	Nrp		**Rp-**		**Rp+**	
	Typical	**Non-Typical**	**Typical**	**Non-Typical**	**Typical**	**Non-Typical**
Hits	.76 (.03)	.70 (.03)	.64 (.04)	.63 (.04)	.92 (.02)	.89 (.03)
A'	.69 (.03)	.84 (.02)	.44 (.06)	.55 (.07)	.76 (.04)	.84 (.03)
B"	.18 (.11)	-.02 (.10)	-.17 (.11)	-.59 (.11)	-.78 (.07)	-.37 (.09)

To assess *A'* for Control group participants, a similar 3 (Item Type: S+, S-, Ns) x 2 (Typicality) RM ANOVA was conducted. This analysis yielded a similar effect of Item Type, $F(2, 64) = 21.68$, $p < .05$, $\eta^2 = .40$, and of Typicality, $F(1, 32) = 5.82$, $p < .05$, $\eta^2 = .15$, but no reliable interaction, $F(2, 64) = .342$, $p > .05$, $\eta^2 = .01$. Planned comparisons demonstrated that differences across Item Type mirrored Control group participants' differences in hit rates: *A'* scores for S- items ($M = 0.49$, $SE = .04$) were not reliably lower than scores for Ns items ($M = .77$, $SE = .03$), $p > .05$, Cohen's $d = .32$, but both were reliably lower than scores for S+ items (*M*

= .80, *SE* = .03), $p < .05$, Cohen's $d = .53$ and .79, respectively (see the right panel of Figure 7). This is consistent with the expectation that Control group participants should not experience RIF, but does not support the hypothesis that restudy should only produce facilitation for restudied items (S+), and no differences across unstudied but related items (S-) and baseline items (Ns).

Finally, bias as *B"* scores was examined across Practice and Control conditions by separate 3 (Item Type) x 2 (Typicality) RM ANOVAs. Participants who received Practice instructions showed a reliable effect of both Item Type, $F(2, 86) = 69.69$, $p < .05$, $\eta^2 = .62$, and Typicality, $F(1, 43) = 5.40$, $p < .05$, $\eta^2 = .11$, on bias scores, as well as a reliable interaction, $F(2, 86) = 10.95$, $p < .05$, $\eta^2 = .23$. Planned comparisons revealed that participants' showed highest response bias for Rp- items relative to both Nrp items, $p < .05$, Cohen's $d = .54$, and Rp+ items, $p < .05$, Cohen's $d = 1.45$, and bias was reliably higher for Nrp items, relative to Rp+ items, $p < .05$, Cohen's $d = 1.01$. In addition, participants showed a bias to respond 'old' to Non-Typical items, relative to Typical items, $p < .05$, Cohen's $d = .35$. Participants under Control conditions exhibited a significant effect of Item Type, $F(2, 64) = 25.22$, $p < .05$, $\eta^2 = .44$. Post-hoc comparison revealed that participants showed differences in response bias across Ns and S- items, $p < .05$, Cohen's $d = .86$ and between Ns and S+ items, $p < .05$, Cohen's $d = 1.25$. However, there were no differences in bias between S- and S+ items, $p > .05$, Cohen's $d = .38$. The effect of Typicality was not reliable, $F(1, 32) = .96$, $p > .05$, $\eta^2 = .03$. However, there was a significant interaction between Item Type and Typicality, $F(2, 64) = 18.66$, $p < .01$, $\eta^2 = .37$, that revealed a stronger response bias to call items 'old' if they were Typical S+ items, relative to Non-Typical S+ items. A full outline of bias scores as a function of Condition, Item Type, and Typicality can be found in Table 2 and Table 3.

Experiment 1b: Method

Participants

Based on the same a priori analysis of Reppa et al. (2017), approximately 60 participants were necessary for Experiment 1, divided between retrieval practice or restudy instructions. In total, 92 participants from Louisiana State University participated in Experiment 1b. Forty participants (M_{age} = 19.7 years, 29 female) completed the experiment under Retrieval Practice conditions, and 33 participants (M_{age} = 19.8 years, 27 female) completed the experiment under Control conditions. One participant from Practice conditions was excluded for chance performance during the retrieval practice phase; data reported for Practice conditions represent the remaining 39 participants. One participant was excluded from analysis under Control conditions due to equipment malfunction; data reported for Control conditions represent the remaining 32 participants. An additional 19 participants completed the experiment with no RIF instructions to provide baseline measure of psychological space; however, 3 participants were excluded from analyses due to equipment malfunction. Baseline data represent the remaining 16 participants. All participants self-reported normal or corrected-to-normal vision, normal color vision, and were native English speakers. Participants engaged in sessions lasting no more than 45 minutes.

Stimuli

Stimuli were identical to those used in Experiment 1a.

Design

The design for Experiment 1b was identical to Experiment 1a with the following exceptions. The dependent variables that were assessed included participants' perceptions of similarity in psychological space. This was accomplished by employing MDS, specifically by having participants arrange objects spatially based on perceived similarity (i.e., the spatial arrangement method, SpAM; Hout et al., 2013). This technique produces item-to-item distances for each participant that can be compared to assess how similarity ratings change as a function of practiced categories. Participants who received no RIF component served as baseline measures of similarity via inter-object distances.

Procedure

The first and second stages of Experiment 1b were identical to Experiment 1a (study phase, recognition practice phase). The final phase consisted of the Spatial Arrangement Method of multidimensional scaling (see Figure 9 for an example). To ensure that scaling distances were not artificially minimized by category membership, participants completed eight SpAM trials, one for each studied category, with category order randomized across participants. On each trial, participants were shown all study items from a single category (e.g., benches, birdhouses, etc.) simultaneously, randomly arranged in rows and intermixed with novel distractors. These items were arranged around an "active arena," into which participants were instructed to drag and position the objects in space such that smaller distance between two objects corresponded with higher perceived similarity. Participants were instructed that they could place objects in overlapping positions or on top of each other, but that this configuration would indicate extremely high levels of similarity (i.e., nearly the same object). Participants were not allowed to advance until all objects were placed within the active arena and were clicked at least one time. To conclude each trial, participants clicked a stop sign in the lower right-hand corner of the screen, and were given the option to move to the next trial, revisit their MDS arrangement, or return all objects to their starting positions and create a new MDS space.

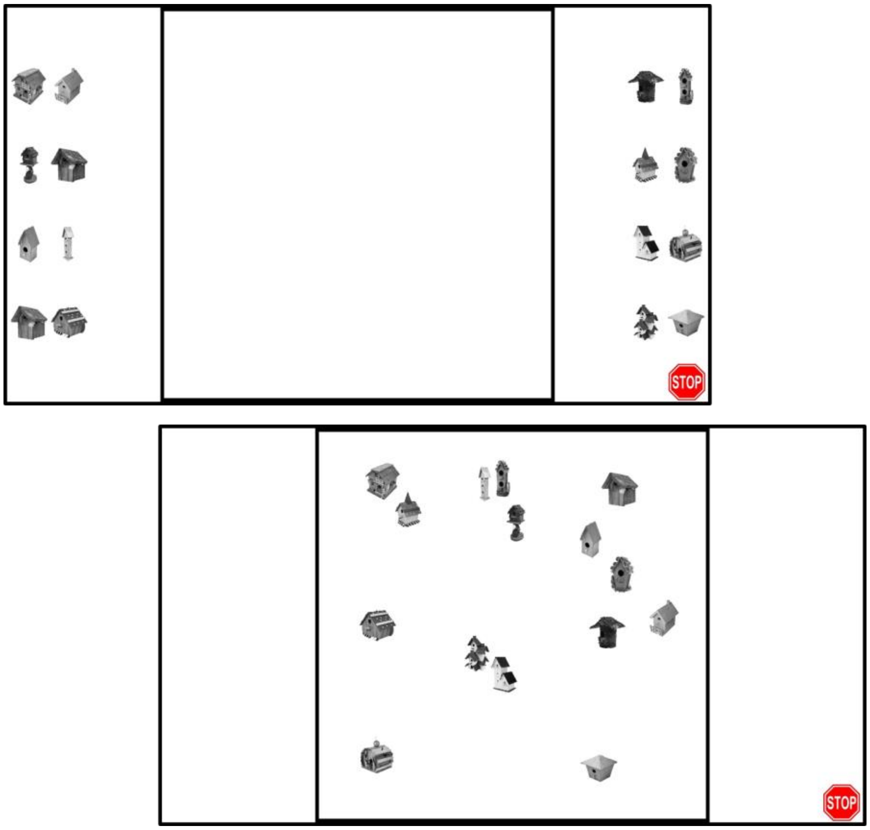

Figure 9. Sample MDS trial of the category Birdhouse. Each trial began with 16 exemplars randomly arranged outside of the active arena (top). To complete the trial, participants clicked and dragged each item into place, with shorter distances indicating greater perceived similarity. To conclude the trial, participants, clicked the stop sign.

Experiment 1b: Results

Recognition Practice

To ensure that participants in the Practice group successfully completed the recognition practice phase, hits (defined as correct recognition of studied objects), false alarms, and *A′* and *B′′* were assessed by separate paired samples *t*-test across typical and non-typical practiced objects. Like Experiment 1a, participants showed numerically higher hit rates for non-typical exemplars (M = .80, SE = .03) than typical studied exemplars (M = .79, SE = .02), but the difference was not statistically reliable, $t(38) = .52$, $p > .05$, Cohen's $d = .08$. However, participants did exhibit a reliably lower false alarm rate for non-typical distractor objects (M =

.32, *SE* = .04), relative to typical distractor objects (*M* = .44, *SE* = .03), *t*(38) = 2.85, *p* < .05, Cohen's *d* = .46. Participants' *A'* scores were not reliably different, with participants exhibiting numerically higher sensitivity for non-typical objects (*M* = .81, *SE* = .03), compared to typical objects (*M* = .67, *SE* = .06), *t*(38) = 1.82, *p* > .05, Cohen's *d* = .29. There was no reliable difference in *B"* scores across non-typical (*M* = -0.31, *SE* = .09) and typical (*M* = -0.45, *SE* = .06) exemplars, *t*(38) = 1.31, *p* > .05, Cohen's *d* = .21. Finally, I expected to find that hits during the recognition practice phase would be significantly above chance for both types of practiced objects (Rugo et al., 2017). Hit rates for non-typical and typical objects were both significantly above chance, *t*s > 11, *p*s < .001, Cohen's *d*s > 1.7. Although participants did not show reliably higher hit rates or sensitivity for non-typical items, false alarm rates again support the hypothesis that non-typical objects are more discriminable than typical objects (Reppa et al., 2017); see Figure 10 for an illustration of false alarm rates and *A'* scores.

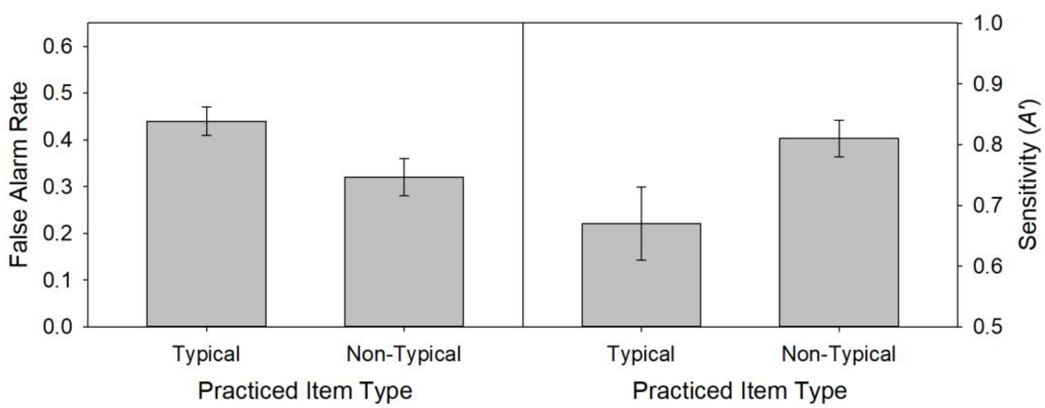

Figure 10. False alarm rates (left) and sensitivity (*A'*) across typical and non-typical items for participants who experienced retrieval practice. Error bars represent SEM.

Multidimensional Scaling Solutions

Using the SpAM method, participants' perceived similarity was assessed by inter-item distances in psychological space. Distances were pooled so that averages were examined for all possible pairs. Average inter-objects distances were assessed by a 3 (Item Type) x 2 (Typicality) RM ANOVA for participants who received Practice instructions. I predicted that average inter-object distances would mirror results from Experiment 1a: For Practice group participants, non-typical objects should have overall higher inter-object distances, relative to typical objects, because they are more memorable and should therefore be more distributed in psychological space (Hout et al., 2014). This prediction was supported by participants' MDS solutions, as there was a significant effect of Typicality, $F(1, 38) = 20.54$, $p < .05$, $\eta^2 = .43$. This effect showed that average distances for Typical items ($M = 428$, $SE = 14$) were reliably smaller than distances for Non-typical items ($M = 458$, $SE = 15$); see Figure 11 for sample psychological spaces produced by two participants using this method. However, I predicted that Practice group participants would show variation depending on the practiced category, with greater inter-object distances for Rp+ items, relative to Nrp items, and smaller inter-object distances for Rp- items, relative to Nrp items. This prediction was not supported, as there was no reliable effect of Item Type, $F(2, 76) = .82$, $p > .05$, $\eta^2 = .02$, and no significant interaction between Item Type and Typicality, $F(2, 76) = .03$, $p > .05$, $\eta^2 < .01$.

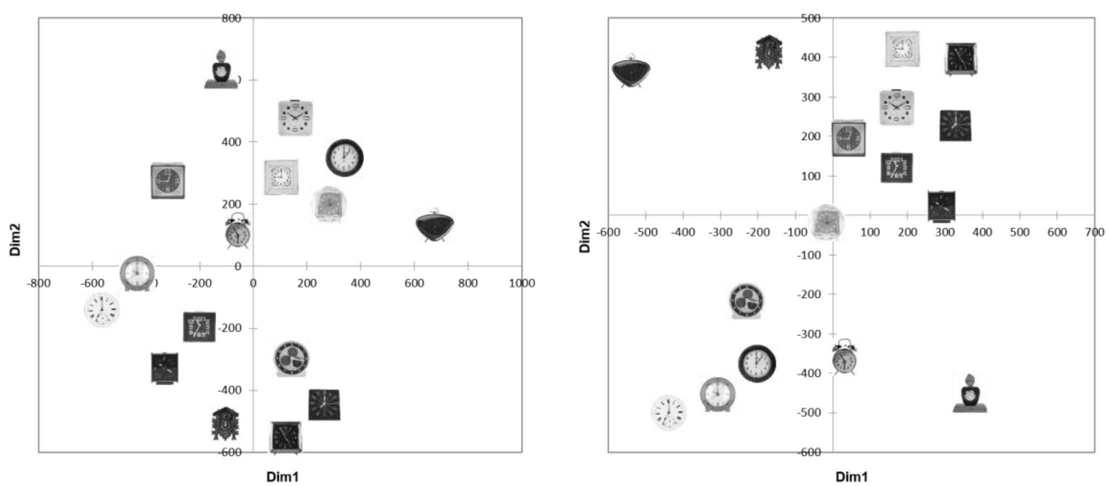

Figure 11. Derived psychological spaces for the category clocks from two participants. Axes represent each participant's distances in psychological space.

To assess whether inter-object distances for Rp- items differed from baseline items (i.e., corresponding items under Control conditions), an additional 2 (Condition: Practice, control) x 3 (Item Type: Nrp/Ns, Rp-/S-, Rp+/S+) x 2 (Typicality) mixed model ANOVA was conducted, with Condition held between subjects. This analysis also revealed no reliable differences across all Item Types, $F(2, 138) = 1.31$, $p > .05$, $\eta^2 = .02$, and no reliable effect of Condition, $F(1, 69) = 2.08$, $p > .05$, $\eta^2 = .03$. The effect of Typicality remained reliable, $F(1, 69) = 66.34$, $p < .05$, $\eta^2 = .48$. Planned comparisons revealed that there was a slightly larger effect of Typicality under Control conditions, although the interaction between Typicality and Condition was not reliable, $F(1, 69) = 3.81$, $p > .05$, $\eta^2 = .03$. The interaction between Item Type and Typicality was not reliable, $F(2, 138) = .09$, $p > .05$, $\eta^2 < .01$, nor were the interactions between Item Type and Condition, $F(2, 138) = 1.14$, $p > .05$, $\eta^2 = .02$, and between Item Type, Typicality, and Condition, $F(2, 138) = .11$, $p > .05$, $\eta^2 < .01$. These results suggest that there were no differences between average inter-object distances between RIF object distances and control object distances. However, it is possible that because participants showed RIF under Restudy conditions in Experiment 1a, the lack of observed differences in psychological space are a function of competition produced under restudy conditions. If RIF occur for visual information when participants simply restudy, as opposed to only following active retrieval practice, this might account for similar distances across Item Types.

To examine average distances across Typicality as a function of RIF Item Type, separate 3-factor (Item Type) RM ANOVAs were conducted (see Figure 12). These analyses revealed no significant differences across Non-typical Nrp items, Rp- items, and Rp+ items , $F(2, 76) = .38$, $p > .05$, $\eta^2 = .01$ and no significant differences across Typical Nrp items, Rp- items, and Rp+ items, $F(2, 76) = .39$, $p > .05$, $\eta^2 = .01$. Although average distances were not statistically reliable, the observed patterns were in the predicted directions, with participants showing slightly larger differences between non-typical Rp+ and Rp- items, relative to the differences between typical Rp+ and Rp- items.

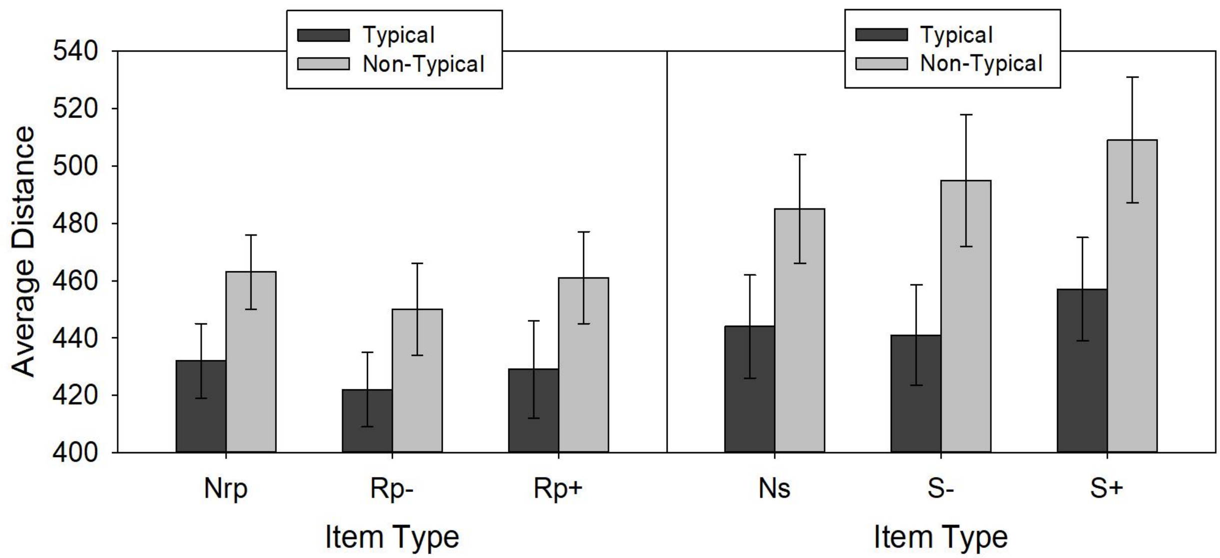

Figure 12. Average inter-item distances for participants who received Practice instructions (left) and Control instructions (right). Error bars represent SEM.

For Control group participants, I predicted that there would only be differences based on Typicality, with greater differences for non-typical objects, relative to typical objects. To investigate inter-object distances, distances were assessed by a 3 (Item Type) x 2 (Typicality) RM ANOVA. This analysis revealed no significant effect of Item Type, $F(2, 62) = 1.43$, $p > .05$, $\eta^2 = .04$, with comparable inter-item distances across Ns items ($M = 464$, $SE = 18$), S- items ($M = 468$, $SE = 18$), and numerically but not reliably higher distances for S+ items ($M = 483$, $SE = 18$). Consistent with my predictions, there was a significant effect of Typicality, $F(1, 31) = 48.51$, $p < .05$, $\eta^2 = .61$. Average inter-item distance for items that were Non-typical ($M = 490$, $SE = 18$) were larger than for items that were Typical ($M = 447$, $SE = 16$), $t(31) = 7.46$, $p < .05$, Cohen's d

= 1.32. The interaction was not reliable, $F(2, 62) = .16$, $p > .05$, $\eta^2 = .01$. Average distances across Typicality and Item Type for Control participants are illustrated in Figure 11.

For participants who received no RIF component, average distances were again examined as a function of Typicality by a paired samples *t* test. This analysis revealed that average distances were once again larger for Non-Typical ($M = 452$, $SE = 21$), relative to Typical ($M = 414$, $SE = 17$) items, $t(15) = 6.93$, $p < .05$, Cohen's $d = 1.73$. To examine whether these distances differed as a function of RIF instructions, an additional 3 (Instruction: RIF, Control, MDS only) x 2 (Typicality) mixed model ANOVA was conducted, with Instruction held between subjects. Like the comparison between Practice and Control conditions, this analysis yielded no reliable effects of Instruction, $F(2, 84) = 1.58$, $p > .05$, $\eta^2 = .04$, a reliable effect of Typicality, $F(2, 84) = 79.16$, $p < .05$, $\eta^2 = .47$, and no reliable interaction, $F(2, 84) = 2.19$, $p > .05$, $\eta^2 = .03$. The results of this analysis suggest that retrieval practice did not significantly affect participants' distribution of objects in psychological space relative to no memory instructions.

Experiment 2. Forgetting Faces

The purpose of Experiment 1 was to investigate the contribution of memorability to recognition-induced forgetting by employing typical and non-typical objects as study material. To further investigate how stronger memories create additional competition between Rp+ and Rp- items, leading to inhibition and/or sharpening of item representations, Experiment 2 aimed to extend the results of Experiment 1 by employing faces, as they are another category of stimuli that naturally range in item strengths. Faces, like everyday objects, contain rich episodic and semantic detail. However, although faces are often touted as a class of visual objects with which human observers are experts, evidence has consistently shown that expertise is limited to faces belonging to highly familiar individuals (e.g., Burton, Wilson, Cowan, & Bruce, 1999). Specifically, there is growing evidence of significant differences between the way familiar faces (i.e., those personally familiar to the observer) and unfamiliar faces (i.e., prior to an experimental context, unknown) are processed (see Johnston & Edmonds, 2009, for a review). These differences can be broadly characterized in that many variables that do not affect familiar face processing (e.g., changes in lighting, expression, or other contextual factors) impair unfamiliar face processing, and there is evidence of qualitative differences in the way these two classifications of faces are perceived.

Many models of face perception differ in how they explain successful face recognition by supposing there are differences in how faces are represented in psychological space. Two common categories of models are *norm-based* models, which describe faces as represented based on their deviation from a prototype (i.e., the norm), and *exemplar-based* models, which describe faces represented in psychological space relative to their perceived distances from other, previously encountered faces. These model categories are not mutually exclusive, but they provide a useful framework for many face perception perspectives. An example is Valentine's (1991) Multidimensional Space (MDS) framework (see also Valentine, 2001; Valentine, Lewis, & Hills, 2016), which proposes that faces are represented in psychological space along specific dimensions useful for discrimination. These dimensions are relative, insomuch as they are determined by exposure to faces over time, and faces are assumed to be normally distributed along any given dimension. The origins of each dimension are assumed to reflect a central tendency, or average face (i.e., the prototype).

Valentine's framework holds that faces represented in psychological space can vary in their distinctiveness, with faces perceived as more similar grouped more tightly. Smaller inter-face distances are associated with greater confusability, contributing to less accurate recognition. This assumption is supported by empirical work showing more distinctive faces are located along the peripheries of psychological space, and more typical faces are clustered nearer the center (Johnston et al., 1997). Distinctiveness, or typicality, of faces shows a direct parallel to object perception; MDS solutions for typical and non-typical objects show similar spread, with non-typical items along the peripheries (Hout et al., 2014). In addition, a key component of Valentine's model is that highly memorable faces are associated with enhanced representation in psychological space by more distinct coding (Valentine, 1991). If RIF serves to sharpen the representations of Rp+ while inhibiting the representations of Rp-, this should be borne out in psychological space for faces: Rp- faces should be perceived as less distinctive (i.e., more similar) than Rp+ and Nrp faces.

Two recent studies have examined recognition-induced forgetting of faces (Ferreira et al., 2014; Rugo, Tamler, Woodman, & Maxcey, 2017). Ferreira et al. (2014) found evidence of RIF of facial features (Experiment 1a) and names (Experiment 2a) when participants studied and practiced faces using cues that highlighted perceptual similarity (e.g., similar hairstyles) or conceptual similarity (e.g., same occupational category). Additionally, Rugo et al. (2017) found that when observers studied Black and White faces and then performed recognition practice on one racial category, RIF was observed, driven primarily by forgetting of White faces. These findings are important for three reasons. First, it suggests that faces are not immune to the effects of RIF, and that inhibitory mechanisms may play a role in retrieving information about identity representations (Ferreira et al., 2014). Second, Rugo et al. (2017) discuss that, overall, White faces were remembered more accurately than Black faces, suggesting in turn that more memorable stimuli show larger magnitudes of forgetting (a similar pattern as more memorable *objects*; see, e.g., Reppa et al., 2017). This result is also informative because observers viewing other-race faces show less spread in psychological space, relative to own-race faces (Papesh & Goldinger, 2010), supporting the notion that own-race faces may be more memorable. Finally, faces are often considered objects of expertise and may be represented in dense neural networks (Anderson, 1974); when one face is retrieved from this network, its activation spreads to other category members, resulting in inhibition (see Rugo et al., 2017). Based on this argument, the more expertise observers have with categories of faces, the denser neural networks become, which should lead to larger observed RIF. However, it is also possible that expertise leads to more effective categorical perception, in that boundaries between category exemplars (in this case, own-race identities) become more pronounced, and the dimensions for within-category comparisons narrow (see Angeli, Davidoff, & Valentine, 2008; Balas, 2012). This allows observers to rely on only the most useful dimensions for discrimination, resulting in identity representations that are more distinct, and theoretically leading to *less* competition during retrieval practice.

Grounded in these findings, there were several hypotheses for Experiment 2 (see Table 4 for a summary of these hypotheses and predicted results). First, participants should show recognition-induced forgetting for Rp- faces (Ferreira et al., 2014; Rugo et al., 2017) in a similar manner as RIF observed for objects (Maxcey & Woodman, 2014; Reppa et al., 2017). Second, if RIF is due to inhibition of Rp-, items, and "sharpening" of Rp+ items, Rp- stimuli should appear more similar in psychological space than Rp+ items and Nrp items when memory is examined using MDS. Third, and consistent with Experiment 1, if own-race faces are more memorable, this arguably creates more competition during retrieval practice in a similar manner as highly memorable objects (Reppa et al., 2017). Faces of a different race than the observer are less distinct, and identities less memorable (Valentine, 1991); this should produce less competition during retrieval practice. Less competition between category members has been associated with less need for inhibitory processes and smaller or absent RIF (Bäuml & Samenieh, 2010; Levy, McVeigh, Marful, & Anderson, 2007; Storm et al., 2007). Considered together, this should result in higher rates of RIF for highly memorable identities (i.e., own-race faces), relative to less memorable identities (i.e., other-race faces). However, other-race faces appear more perceptually similar to observers and are grouped more closely together in psychological space (Papesh & Goldinger, 2010), similar to the way that typical objects share additional features and strong category associations (Hout et al., 2014). If interference dependence relies more on baseline memory strength (Reppa et al., 2017; Spitzer, 2014), the magnitude of RIF should be weighted more heavily by the memorial strength of faces than by perceptual similarity, with own-race faces producing larger RIF. On the other hand, if perceptual similarity creates additional competition between identities, other-race faces should show larger degrees of RIF.

Still, perceptual similarity should affect distributions in psychological space, with other-race (i.e., highly similar) faces associated with smaller inter-face distances (Papesh & Goldinger, 2010; Valentine, 1991). If RIF serves to sharpen representations of Rp+ items by inhibiting items that compete during retrieval practice, this should result in smaller distances for Rp- faces relative to Nrp faces, regardless of the practiced category. If memory strength produces additional competition and necessitates additional inhibition, the magnitude of inter-face distances between Rp+ and Rp- identities should be larger for own-race faces than for other-race faces. Alternatively, if shared features produce additional competition, differences between Rp+ and Rp- identities should be larger for other-race faces than for own-race faces.

Table 4. Hypotheses and Predicted Results as applied to Experiment 2.

Hypothesis	**Predicted Results**
1. Stronger memories produce additional competition when unpracticed.	Own-race Rp- faces, which are highly memorable, should show RIF.
2. Perceptually similar items produce additional competition when unpracticed.	Other-race Rp- faces, which are less discriminable, should show RIF.
3. Perceived similarity in psychological space will change as a function of RIF.	Rp- faces will show smaller distances (i.e., appear more similar) relative to Nrp items.

Experiment 2a: Method

Participants

Based on an a priori analysis of Rugo et al. (2017), who found RIF across own-and other-race faces, approximately 60 participants were necessary for Experiment 2a. This estimate was based on mean differences across Rp+ and Rp- faces, with significant RIF observed, with Cohen's *d* = .68, *α* = 0.05, and power held at 0.95. In total, 116 participants from Louisiana State University completed the experiment in exchange for partial course credit. Seventy-two participants (M_{age} = 19.5 years, 57 female) received Retrieval Practice instructions, and the remaining 43 participants (M_{age} = 20.6 years, 33 female) received Control instructions. All participants self-reported normal or corrected-to-normal vision, normal color vision, and were native English speakers. Participants engaged in individual sessions lasting no more than 45 minutes.

To ensure that other-race effects were appropriately manipulated, data was only collected and analyzed from participants who self-identified as Caucasian/White or Asian-American/Asian. These racial/ethnic groups were chosen based on student demographics of the available participant pool at Louisiana State University. Fall 2018 estimates of student body by race/ethnicity demonstrate approximately 4% of LSU students self-identify as Asian/Asian-American, while approximately 67% of LSU students self-identify as Caucasian/White (LSU Fall Facts, 2018). These estimates parallel demographics of the state of Louisiana, with 62.6% Caucasian and only 1.7% Asian residents (Louisiana Population, 2019). These demographics were considered because other-race deficits differ based on contact observers have with other races (see Hugenberg, Young, Bernstein, & Sacco, 2010). Although Caucasian/White and African-American/Black are often used in investigations of other-race effects, considered with the fact that approximately 12% of the LSU student body and approximately 32.2% of the population of the state of Louisiana identifies as African-American/Black, Asian identities were chosen to maximize observed other-race effects.

One participant was excluded from analysis for self-identifying as African-American/Black. From the remaining participants, 5 who received Practice instructions and 5 who received Control instructions self-identified as Asian-American/Asian, while the remaining 105 participants self-identified as Caucasian. Thirty-seven participants completed recognition practice with own-race faces (White faces for White participants, and Asian faces for Asian participants), 35 completed recognition practice with other-race faces (Asian faces for White participants, and White faces for Asian participants), 21 experienced restudy with own-race faces, and 22 experienced restudy with other-race faces.

Stimuli

A total of 146 possible faces were used in the current experiment. Of these, faces were divided into two categories based on distinctiveness to Caucasian participants, own-race (White) faces, and other-race (Asian) faces (although these labels were modified for participants who identified as Asian-American/Asian). These race categories were chosen based on the racial demographics of available participants as outlined above, with all participants recruited from LSU students. There were 89 unique White identities, and 59 unique Asian identities. Images were sourced from the Chicago Face Database (Ma, Correll, & Whittenbring, 2015), which contains high-quality, full-color images of White, Black, and Asian individuals. All faces used were females who displayed neutral expressions, and each face image was sized to fit within 575 x 600 pixels.

Design

Participants were divided based on Recognition Practice or Control instructions. The design for the Practice group was again repeated measures, with participants experiencing Race within-subjects (own-race, other-race), and tested on three Item Types (Rp+, Rp -, and Nrp). Practice group participants practiced retrieving half of the studied faces from one studied race. For the Control group, participants restudied faces but experienced no recognition practice. The dependent variables that were examined included Practice group participants' hit rates for practiced (Rp+), unpracticed but related (Rp-), and baseline (Nrp) faces, Control group participants' hit rates for restudied (S+), not restudied but related (S-), and nonstudied (Ns) faces, false alarms, and overall sensitivity as *A'* scores.

Procedure

There were three phases for the Practice group, including study, recognition practice, and test, and three phases for the Control group, including study, restudy, and test. See Figure 13 for an illustration of the study and recognition practice phases, and Figure 14 for an illustration of the recognition test phase. Each phase is described in detail below.

Study Phase. During the study phase, Practice and Control participants began each trial with a 500 ms fixation cross. They then viewed individual faces for 5 seconds each. Unlike Experiment 1, because unfamiliar face memory is relatively poor relative to familiar face memory (Hill & Bruce, 1996; O'Toole et al., 1998; see also Burton, 2013), participants were encouraged to learn faces. Participants studied 20 own-race and 20 other-race faces, for a total of 40 faces. Each face was followed by a 500 ms inter-trial interval. Upon completion of the study phase, participants experienced a 5-minute break, during which they performed the same visual distractor task described in Experiment 1, a simple visual search for Waldo in digitally adapted *Where's Waldo?* images (Maxcey, 2016; Rugo et al., 2017).

Recognition Practice Phase. Practice group participants experienced a second phase during which they performed a recognition memory test on half of the studied faces from one studied race, counterbalanced across participants. On each trial, a single face was presented. Participants indicated whether the face was "old" (i.e., previously studied) or "new" (i.e., novel to the experiment). Participants completed two blocks of 20 recognition practice trials, during which they were presented with 10 studied and 10 novel but category-consistent faces (a total of 20 unique identities). Each practiced face appeared twice (Rugo et al., 2017), mixed with an

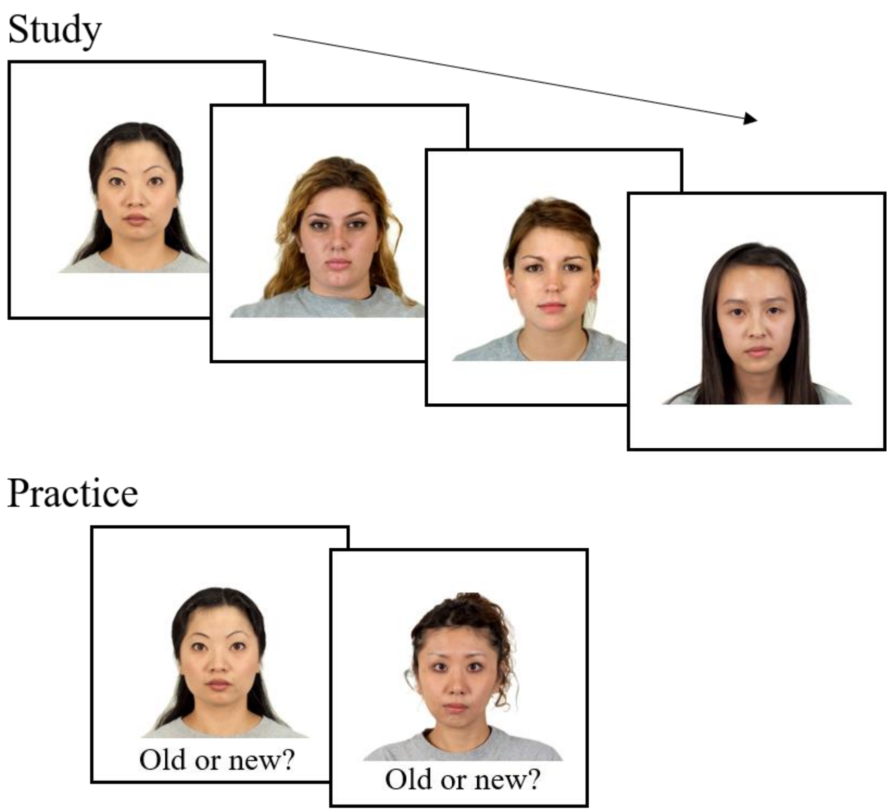

Figure 13. A schematic of the study phase and the retrieval practice phase. Participants studied Asian and White faces, and some participants completed a recognition practice phase on one category of faces only.

equal number of race-consistent distractors. Participants who received Control instructions were re-exposed to half of the studied faces from one studied race twice, along with an equal number of exposures to novel faces, with no recognition component. Between iterations of the practice block, and upon completion of the second phase, participants again performed a visual distractor task of searching *Where's Waldo?* scenes.

Test Phase. All participants completed a test phase, during which they made old/new recognition judgments about studied and unstudied faces. All 40 faces from the study phase were randomly intermixed with the same number of novel faces drawn from each category (i.e., 20 novel White faces and 20 novel Asian faces), for a total of 80 recognition decisions. Practice

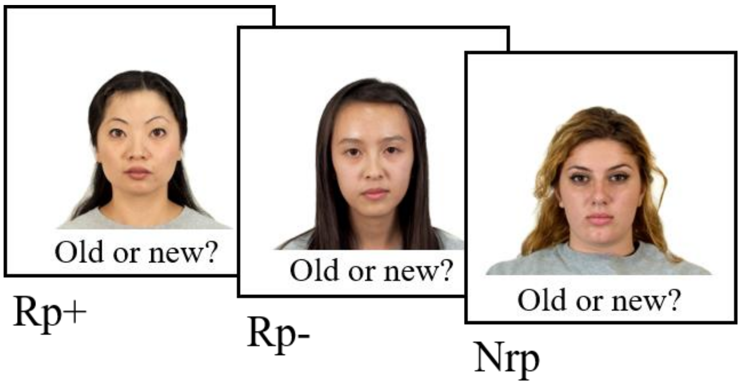

Figure 14. A schematic of the test phase. Participants were tested on studied and practiced faces (Rp+), faces from the practiced category that were studied but not practiced (Rp-), and faces that were studied but not practiced (Nrp).

group participants experienced three types of studied faces: Rp+ faces (practiced), Rp- faces (unpracticed, related), and Nrp faces (unpracticed, unrelated). Control group participants also experienced three types of faces: S+ faces (restudied), S- faces (not restudied, related), and Ns faces (not restudied, not related). Participants indicated whether a face was previously seen at any point during the experiment (i.e., old), or if the face was novel (i.e., new).

Experiment 2a: Results

Recognition Practice

To ensure that participants in the Practice group successfully completed the recognition practice phase, hits (defined as correct recognition of previously studied faces), false alarms, and A' and B'' for studied items were assessed by separate independent samples *t* tests comparing performance across own-race and other-race practiced faces. I expected to find higher hit rates and sensitivity (as A' scores) for participants who performed retrieval practice with own-race faces, which would indicate that White faces are more memorable to White participants and Asian faces more memorable to Asian participants. No patterns of data differed when analyzing performance for only White participants, so full data sets are reported below.

Participants' hit rates across Own- and Other-race practice did not reliably differ, with participants exhibiting numerically higher hits for Own-race practice faces ($M = .86$, $SE = .02$) than Other-race practice faces ($M = .81$, $SE = .02$), $t(70) = 1.58$, $p > .05$, Cohen's $d = .37$. However, participants' false alarm rates were higher for Other-race distractor faces ($M = .20$, $SE = .02$), relative to Own-race distractor faces ($M = .11$, $SE = .02$), $t(70) = 3.58$, $p < .05$, Cohen's $d = .85$. Similarly, participants who practiced Other-race faces showed lower *A'* scores ($M = .88$, $SE = .01$), compared to participants who practiced Own-race faces ($M = .92$, $SE = .01$), $t(70) = 3.3$, $p < .05$, Cohen's $d = .78$. The pattern of false alarm rates and *A'* scores are illustrated in Figure 15. There were no reliable differences in *B''* scores, $t(70) = 1.12$, $p > .05$, Cohen's $d = .26$, nor did *B''* scores significantly differ from zero, $t(71) = 21.74$, $p < .05$. Finally, recognition practice hits were significantly above chance for Own-Race faces, $t(34) = 15.99$, $p < .05$, Cohen's $d = 2.70$, and for Other-Race faces, $t(36) = 15.06$, $p > .05$, Cohen's $d = 2.48$, which suggests participants engaged in successful retrieval across both practiced races, but supports the hypothesis own-race identities were more memorable to participants.

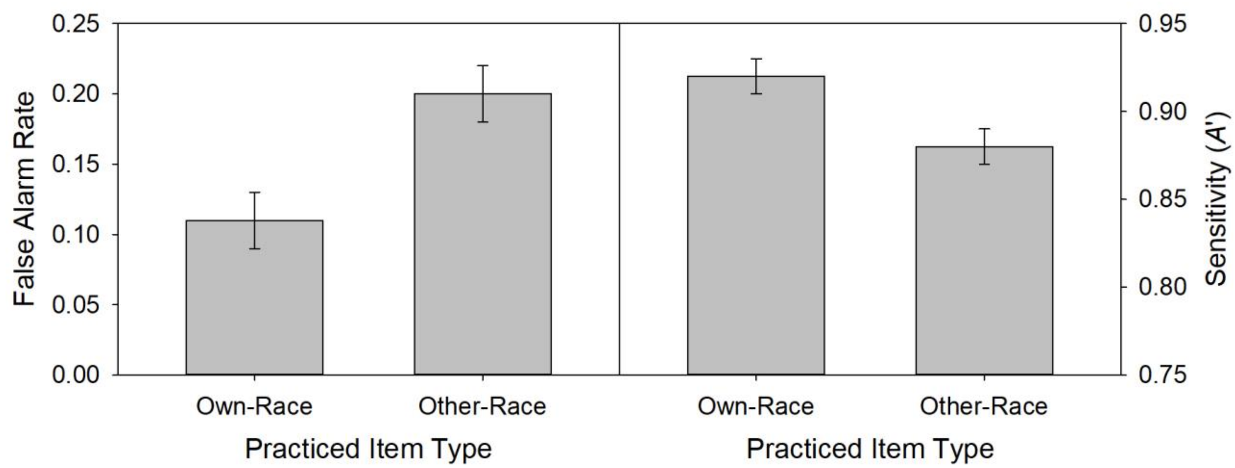

Figure 15. False alarm rates (left) and sensitivity (*A'*, right) across own- and other-race faces for participants who experienced retrieval practice. Error bars represent SEM.

Recognition Test

To assess participants' memory performance, hits (again defined as correct recognition of studied faces), false alarms to novel race-consistent distractor faces, sensitivity as *A'* scores, and bias as *B''* scores was examined. A summary of these metrics is available in Table 4 for participants who received Practice instructions, and Table 5 for participants who received Control instructions.

Table 5. Hits, sensitivity as *A′* scores, and bias as *B″* scores for participants who received Practice instructions.

	Nrp		**Rp-**		**Rp+**	
	Own-Race	**Other-Race**	**Own-Race**	**Other-Race**	**Own-Race**	**Other-Race**
Hits	.59 (.03)	.57 (.03)	.55 (.04)	.52 (.03)	.84 (.02)	.84 (.03)
A'	.83 (.01)	.83 (.01)	.84 (.02)	.46 (.23)	.94 (.01)	.88 (.01)
B"	.36 (.08)	.79 (.05)	.62 (.09)	.70 (.05)	.17 (.12)	.06 (.11)

First, to assess whether all participants show RIF regardless of the study material, hits (again defined as correct recognition of studied faces) for participants in the Practice group were assessed by separate 3 (Item Type: Rp +, Rp -, Nrp) factor RM ANOVAs. I predicted that participants would show significant RIF, with better memory for studied faces (Rp+), relative to baseline faces (Nrp), and worse memory for related but unstudied faces (Rp-), relative to baseline faces (Nrp). This prediction is consistent with Experiment 1a, and with previous evidence that faces (Ferriera et al., 2014; Rugo et al., 2017) and highly memorable visual objects (Reppa et al., 2017) are not immune to recognition-induced forgetting. Participants' hit rates showed a reliable effect of Item Type, $F(2, 142) = 105.8$, $p < .05$, $\eta^2 = .60$. Planned comparisons revealed that hit rates for Rp+ items ($M = .84$, $SE = .02$) was significantly higher than hit rates for Rp- items (*M* = .53, *SE* = .03), $p < .05$, Cohen's $d = 1.59$, and Nrp items (*M* = .58, *SE* = .02), $p < .05$, Cohen's $d = 1.44$, while Rp- and Nrp items did not reliably differ, $p > .05$, Cohen's d = .23 (see the left panel of Figure 16). To investigate differences between Rp+, Rp-, and Nrp items as a function of the race participants practiced, hits were also assessed by separate 3 (Item Type) x 2 (Practiced Race: Own, other) mixed model ANOVAs, with Practiced Race held between subjects. Results of this analysis again yielded a reliable effect of Item Type on hit rates, $F(2, 140) = 104.39$, $p < .05$, $\eta^2 = .60$. There was no reliable effect of Practiced Race, $F(1, 70) = .26$, $p > .05$, $\eta^2 < .01$, and no reliable interaction, $F(2, 140) = .35$, $p > .05$, $\eta^2 < .01$, illustrated in the left panel of Figure 19.

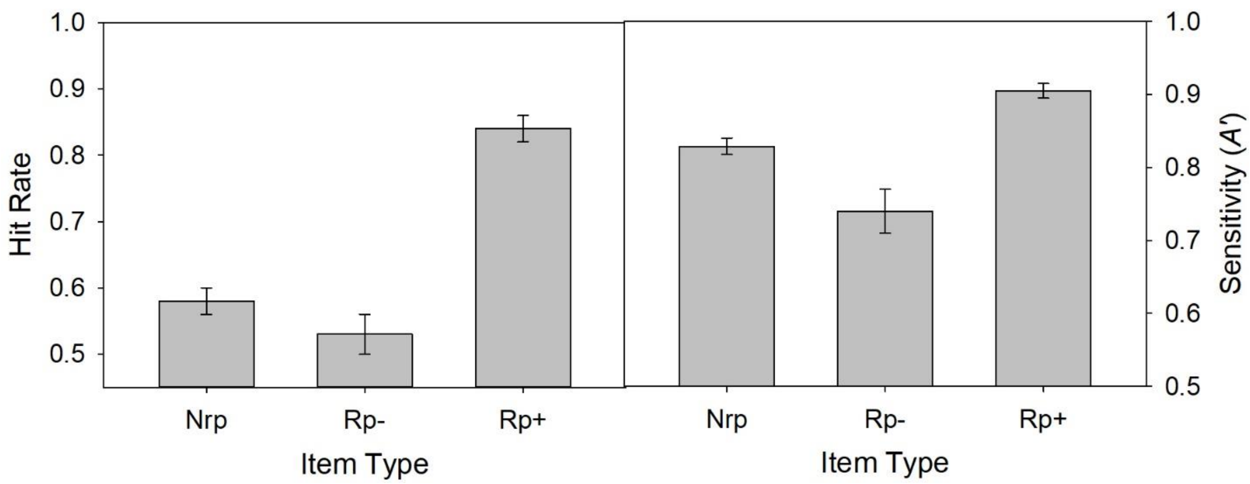

Figure 16. Hit rates (left) and sensitivity (*A'*, right) across Nrp items, Rp- items, and Rp+ items for participants who experienced retrieval practice. Error bars represent SEM.

Control group participants' hits were also assessed by a 3 (Item Type: S+, S-, Ns) factor RM ANOVA. Like Experiment 1a, I predicted that participants should show higher hit rates for restudied (S+) identities, but no difference in hits across related but not restudied (S-) and nonstudied (Ns) identities. For hit rates, this analysis yielded a significant effect of Item Type, $F(2, 84) = 54.46$, $p < .05$, $\eta^2 = .57$. Planned comparisons demonstrated that participants' hit rates for S+ items (*M* = .94, *SE* = .02) was significantly higher than hits for S- (*M* = .64, *SE* = .03) and

Ns (M = .68, SE = .03) faces, $p < .05$, while S- and Ns faces did not reliably differ, $p > .05$, see the left panel of Figure 17. To investigate whether the benefits of restudy were larger for other-race faces, which are more difficult to remember (see Meissner & Brigham, 2001) and to discriminate (Papesh & Goldinger, 2014; Valentine, 1991), hits and *A'* scores were also examined by a 3 (Item Type) x 2 (Practiced Race) mixed model ANOVA, with Practiced Race held between subjects. There was a significant effect of Item Type, $F(2, 82) = 57.33$, $p < .05$, $\eta^2 = .56$, and no effect of Practiced Race, $F(1, 41) = .34$, $p > .05$, $\eta^2 = .01$. The interaction was not reliable, $F(2, 82) = 2.27$, $p > .05$, $\eta^2 = .02$, illustrated in Figure 20. Planned comparisons revealed that across both Practiced Races, there were reliable differences between hit rates for Ns and S- faces, both relative to S+ faces, $ps < .01$. Ns and S- faces did not reliably differ, $p > .05$.

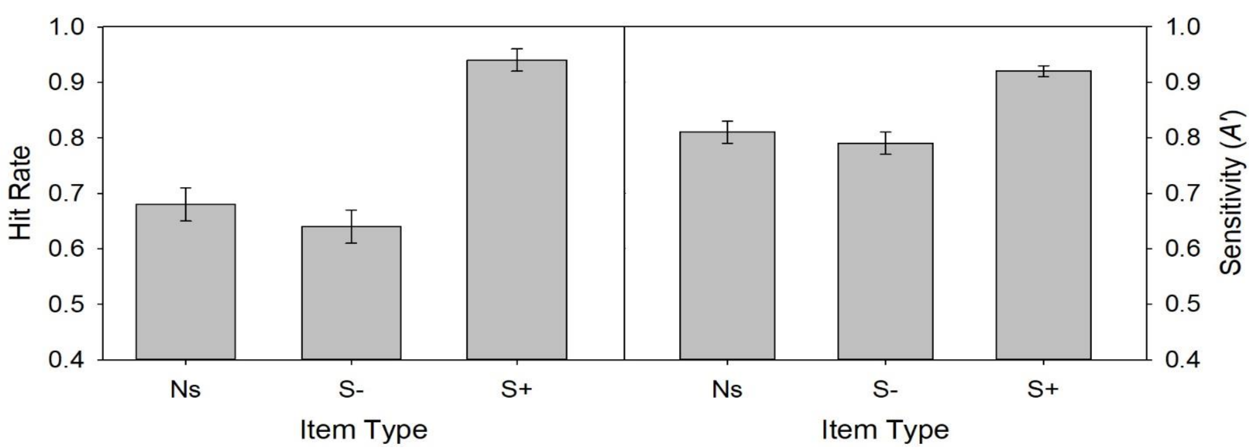

Figure 17. Hit rates (left) and sensitivity (*A'*, right) across Ns, S-, and S+ items, for participants who experienced restudy conditions. Error bars represent SEM.

Hits across Practice and Control conditions were also assessed as a function of Item Type by a 2 (Instruction: Practice, Control) x 3 (Item Type: Rp+/S+, Rp-/S-, Nrp/Ns) mixed model ANOVA, with Instruction held between subjects. This analysis revealed that participants who experienced Restudy conditions had slightly higher hit rates overall (M = .75, SE = .03) than participants who experienced Practice conditions (M = .65, SE = .02), $F(1, 113) = 12.40$, $p < .05$, $\eta^2 = .10$. Although I predicted the overall effect of Item Type for Practice conditions, it is possible that because participants under Practice conditions experienced response-terminated retrieval practice, they did not spend as much time encoding identities as participants under Restudy conditions. This is supported by participants' average response times during retrieval practice (M = 1500 ms, SE = 56 ms), which is significantly less time than the amount participants had to restudy identities (5 seconds), $t(71) = 26.88$, $p < .05$, Cohen's $d = 3.17$.

Next, false alarms (defined as incorrect identification of novel faces as previously studied) for own- and other-race identities were assessed by a paired samples *t*-test comparing false alarms across Race. I predicted that participants should differ in false alarms across Race, with higher false alarm rates for other-race faces. False alarm rates did show a reliable difference, $t(71) = 6.11$, $p < .05$, Cohen's $d = .72$, with higher false alarm rates for Other-race faces (M = .17, SE = .02) compared to Own-race faces (M = .08, SE = .01), illustrated in Figure 18. This finding is consistent with evidence that other-race faces are more difficult to

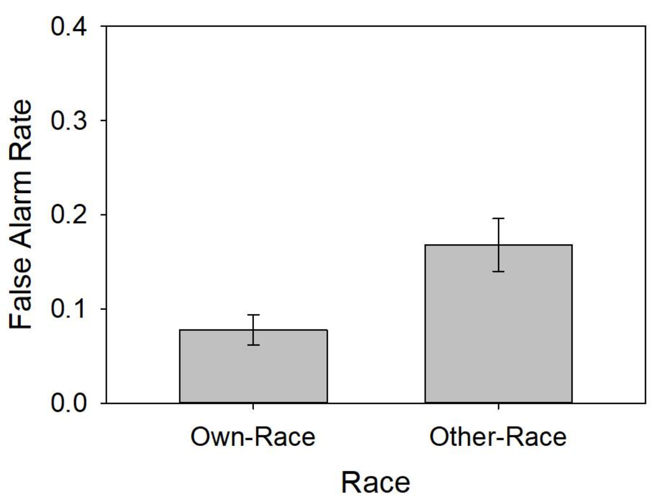

Figure 18. False alarm rates across own- and other-race distractors during the recognition memory test. Error bars represent SEM.

discriminate than own-race faces (Papesh & Goldinger, 2010; Valentine, 1991), and provides additional evidence that other-race faces are less memorable and may not produce significant competition between Rp+ and Rp- items during retrieval practice.

A' scores were also examined for Practice and Control participants as a function of the studied Item Type. For Practice group participants, a 3 (Item Type) factor, RM ANOVA showed a similar pattern as hit rates, with a reliable effect of Item Type, $F(2, 142) = 19.84$, $p < .05$, $\eta^2 = .22$. Planned comparisons revealed *A'* scores (see the right panel of Figure 19) were reliably lower for Rp- items ($M = .74$, $SE = .03$), relative to Nrp items ($M = .83$, $SE = .01$), $p < .05$, Cohen's $d =$, and both were reliably lower than Rp+ items ($M = .91$, $SE = .01$), all *p*s < .05. These findings do not fully replicate RIF using own- and other-race faces as stimuli. However, these data suggest there is some benefit for items that receive retrieval practice, and provide some support for RIF effects observed using faces as stimuli. For Control group participants, a

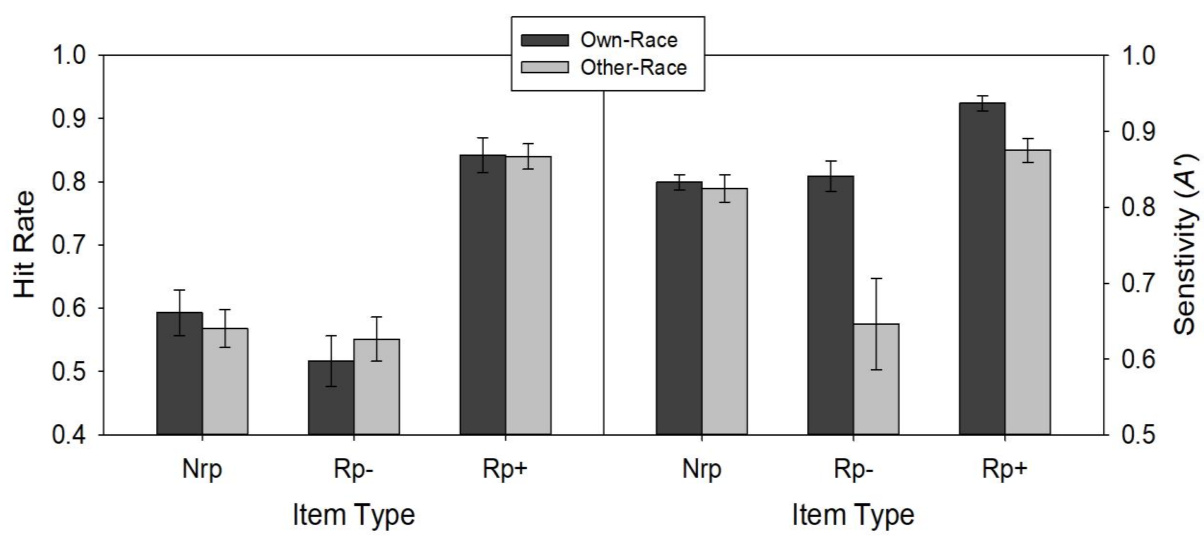

Figure 19. Hit rates (left) and sensitivity (*A'*, right) as a function of Practiced Race and Item Type for participants who experienced retrieval practice. Error bars represent SEM.

similar 3 (Item Type) factor, RM ANOVA revealed participants had reliably higher *A'* scores for S+ faces (*M* = .92, *SE* = .01), and comparable *A'* scores for Ns (*M* = .81, *SE* = .02) and S- (*M* = .79, *SE* = .02) faces, $F(2, 40) = 44.58$, $p < .05$, $\eta^2 = .52$, see the right panel of Figure 20. Combined, these patterns suggest that restudy does produce facilitation for S+ identities, but does not produce competition between Ns and S- identities.

To investigate whether differences between Rp+, Rp-, and Nrp items differed as a function of the practiced race, *A'* scores for Practice and Control group participants were also examined by a 3 (Item Type) x 2 (Practiced Race) mixed model ANOVA, with Practiced Race held between subjects. For Practice group participants, results revealed a reliable effect of Item Type on *A*' scores, $F(2, 140) = 21.10$, $p < .05$, $\eta^2 = .21$. There was also a reliable effect of Practiced Race, with participants showing higher *A'* scores for own-race faces (*M* = .87, *SE* = .02) than scores for other-race faces (*M* = .78, *SE* = .03), $F(1, 70) = 8.40$, $p < .05$, $\eta^2 = .11$. Finally, there was a reliable interaction between Item Type and Practiced Race, $F(2, 140) = 7.33$, $p < .05$, $\eta^2 = .08$, illustrated in the right panel of Figure 19. The results of this interaction revealed

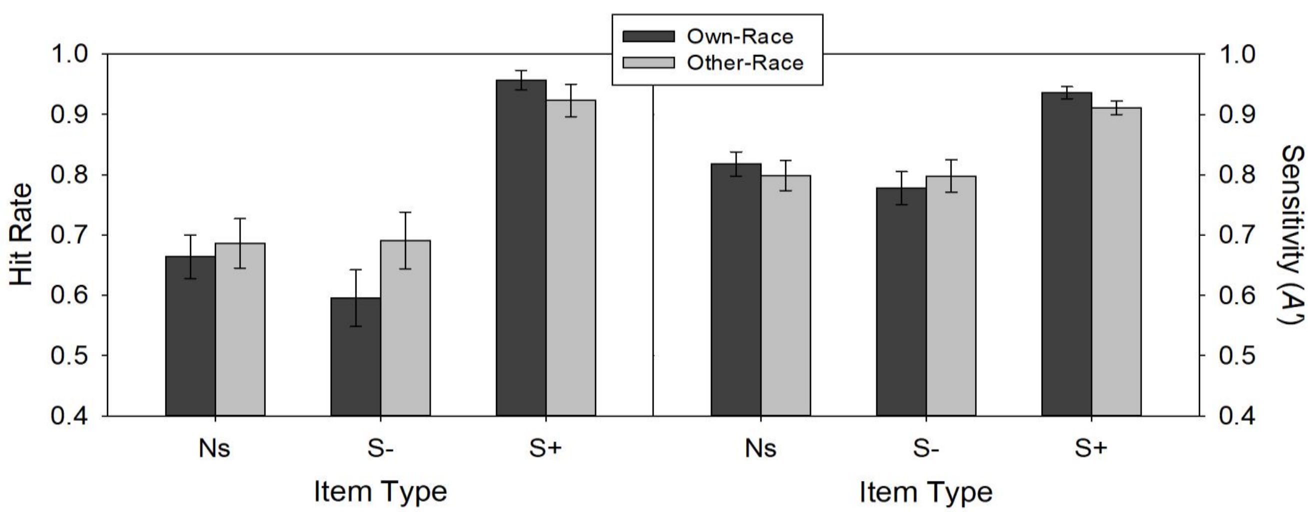

Figure 20. Hit rates (left) and sensitivity (*A'*, right) as a function of Practiced Race and Item Type for participants who experienced restudy. Error bars represent SEM.

that sensitivity for Rp- faces was significantly lower only when the practiced category was Other-race faces. For Control group participants, *A'* scores showed a similar pattern, with a significant effect of Item Type, $F(2, 84) = 45.10$, $p < .05$, $\eta^2 = .52$, but no effect of Practiced Race, $F(1, 41) < 1$, $p > .05$, $\eta^2 < .01$, and no reliable interaction, $F(2, 84) = 1.27$, $p > .05$, $\eta^2 = .02$. However, planned comparisons again revealed that participants showed a significant difference in *A'* scores across Ns and S- faces, compared to S+ faces, $p < .05$, while Ns and S- faces did not reliably differ across either Practiced Race, $p > .05$. The pattern of hit rates *A'* scores supports the prediction that restudy produces facilitation for restudied items only, and does not produce reliable RIF effects.

Finally, participants' bias as *B"* for Practice and Control groups was examined by a 3 (Item Type) x 2 (Practiced Race) mixed model ANOVA, with Practiced Race held between subjects. For Practice conditions, this analysis revealed a significant effect of Item Type, $F(2, 140) = 44.82$, $p < .05$, $\eta^2 = .36$, no significant effect of Practiced Race, $F(1, 70) = 1.49$, $p > .05$, $\eta^2 = .02$, and a reliable interaction, $F(2, 140) = 10.24$, $p < .05$, $\eta^2 = .08$. This pattern, illustrated in Table 4, suggested that participants showed higher rates of response bias when responding to Nrp and Rp- items, as both were reliably higher than bias for Rp+ items, $p < .05$, Cohen's $d = .70$ and 1.01, respectively, while Nrp and Rp- items did not reliably differ, $p > .05$, Cohen's $d = .18$. Similarly, participants under Control conditions showed a reliable effect of Item Type, $F(2, 82) = 69.89$, $p < .05$, $\eta^2 = .61$, no reliable effect of Practiced Race, $F(1, 41) = 1.21$, $p > .05$, $\eta^2 = .03$, and no reliable interaction, $F(2, 82) = 3.10$, $p > .05$, $\eta^2 = .03$. Overall, the pattern of results was very similar, with participants showing higher bias scores for Ns and S- faces, relative to S+ faces, $p < .05$, Cohen's $d = 1.50$ and 1.42, respectively, while Ns and S- faces did not reliably differ, $p > .05$, Cohen's $d = .22$.

Table 6. Hits, sensitivity as *A′* scores, and bias as *B″* scores for participants who received Control instructions.

	Nrp		**Rp-**		**Rp+**	
	Own-Race	**Other-Race**	**Own-Race**	**Other-Race**	**Own-Race**	**Other-Race**
Hits	.66 (.03)	.69 (.04)	.60 (.04)	.69 (.04)	.96 (.01)	.92 (.02)
A’	.82 (.02)	.80 (.02)	.78 (.03)	.80 (.02)	.94 (.01)	.91 (.01)
B”	.36 (.08)	.79 (.05)	.62 (.09)	.70 (.05)	.17 (.12)	.06 (.11)

Considered altogether, these findings do not support the hypothesis that items that are associated with a stronger memory trace (i.e., own-race identities) produce a larger magnitude RIF, as the difference between hit rates Rp+ items, and Rp- and Nrp items, did not differ as a function of Practiced Race. However, it is possible that because unfamiliar face memory is poor even under optimal viewing conditions (see Johnston & Edmonds, 2009), these weak memory traces do not produce competition between Rp+ and Rp- items during practice.

Experiment 2b: Method

Participants

Based on the same a priori analysis of Rugo et al. (2017), approximately 60 participants were necessary for Experiment 2b. In total, 133 participants from Louisiana State University completed the experiment in exchange for partial course credit. Fifty-seven participants (M_{age} = 20.6 years, 30 female) received Retrieval Practice instructions, and the remaining 32 participants (M_{age} = 20.5 years, 24 female) received Control instructions. An additional 16 participants completed the task under no RIF instructions (i.e., MDS only with own- and other-race identities) to provide baseline measure of psychological space. All participants self-reported normal or corrected-to-normal vision, normal color vision, and were native English speakers. Participants engaged in individual sessions lasting no more than 45 minutes.

To ensure that other-race effects were appropriately manipulated, data was only collected and analyzed from participants who self-identified as Caucasian/White or Asian-American/Asian. Two participants were excluded from analysis in the Practice group for self-identifying as African-American/Black, and two participants were excluded for failing to follow experimental protocols. The results reported below include data from the remaining 53 participants. Twenty-nine participants completed recognition practice with own-race faces (White faces for White participants, and Asian faces for Asian participants), 24 completed recognition practice with other-race faces (Asian faces for White participants, and White faces for Asian participants), 17 experienced restudy with own-race faces, and 15 experienced restudy with other-race faces.

Stimuli

Stimuli were identical to those used in Experiment 2a with the following exception. For MDS decisions, faces were scaled smaller than study faces to fit within approximately 200 x 140 pixels. This was approximately 1/3 the size of studied faces, but was large enough to provide detailed feature information during sorting decisions.

Design

The design for Experiment 2b was identical to Experiment 2a. However, the dependent variables that were assessed included participants' perceptions of similarity in psychological space. This was accomplished by employing MDS, specifically by having participants arrange faces spatially based on perceived similarity. This technique produces item-to-item distances for each participant that can be compared to assess how similarity ratings change as a function of the practiced (or restudied) category. In addition, hit rates, false alarm rates, *A'* scores, and *B''* scores were assessed during retrieval practice for participants who received Practice instructions.

Procedure

The first and second stages of Experiment 2b were identical to Experiment 2a (study phase, recognition practice phase). The final phase consisted of the Spatial Arrangement Method (SpAM) of multidimensional scaling identical to Experiment 1b (see Figure 21 for an example), with the following exceptions. To ensure that inter-item distances were not artificially minimized by participants considering race in their scaling decisions, participants completed multiple SpAM trials. On each trial, participants saw studied faces from a single race (i.e., all 20 Asian faces or all 20 White faces). Participants were instructed that their task was to arrange faces in the "active

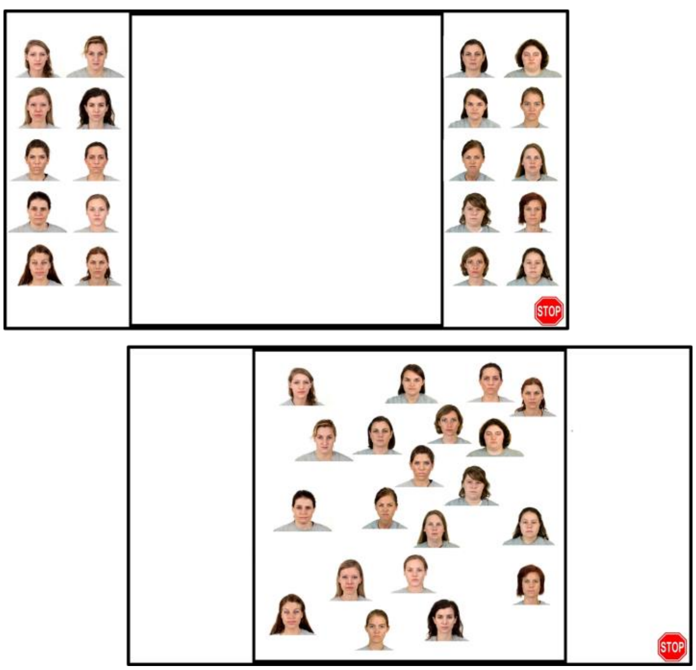

Figure 21. Sample MDS schematic. Each trial began with 20 faces flanking either side of the active arena (top). Participants clicked and dragged faces into the active arena, and concluded the trial by clicking the stop sign in the lower right corner. Note faces are scaled slightly larger here than they appeared in actual trials for ease of viewing.

arena" in a way that shorter distances represented more similar identities, and larger distances represented more dissimilar identities. Participants were instructed that although faces could overlap, this would indicate extremely high levels of perceived similarity (i.e., the same identity). All participants completed scaling decisions for both Asian and White faces, with scaling order counterbalanced across participants.

Experiment 2b: Results

Recognition Practice

To ensure that participants in the Practice group successfully completed the recognition practice phase, hits (defined as correct recognition of previously studied faces), false alarms, and *A′* and *B″* for studied items were assessed by separate independent samples *t* tests comparing performance across own-race and other-race faces. Like Experiment 2a, I expected to find higher hit rates and *A′* scores for participants who practiced own-race faces, which would indicate that White faces are more memorable to White participants and Asian faces more memorable to Asian participants.

Participants' hit rates across Own- and Other-race practice reliably differed, with participants exhibiting higher hit rates when practicing Own-race faces (M = .86, SE = .02), relative to Other-race faces (M = .79, SE = .02), $t(51) = 2.29$, $p < .05$, Cohen's d = .63. Participants' false alarm rates were also higher for Other-race distractor faces (M = .18, SE = .02), relative to Own-race distractor faces (M = .09, SE = .01), $t(51) = 3.57$, $p < .05$, Cohen's d = .99. Similarly, participants who practiced Other-race faces showed lower *A'* scores (M = .87, SE = .02), compared to participants who practiced Own-race faces (M = .93, SE = .01), $t(51) = 51$, $p < .05$, Cohen's d = .94; see Figure 22 for an illustration of false alarm rates and *A'* scores. There were no reliable differences in *B"* scores, $t(51) = .23$, $p > .05$, Cohen's d = .01, nor did *B"* scores significantly differ from zero, $t(52) = 1.60$, $p > .05$. Finally, recognition practice hits were significantly above chance for both Practiced Race categories (Rugo et al., 2017), $t(52) = 20.10$, $p < .05$, Cohen's $d = 2.76$, which suggests successful retrieval.

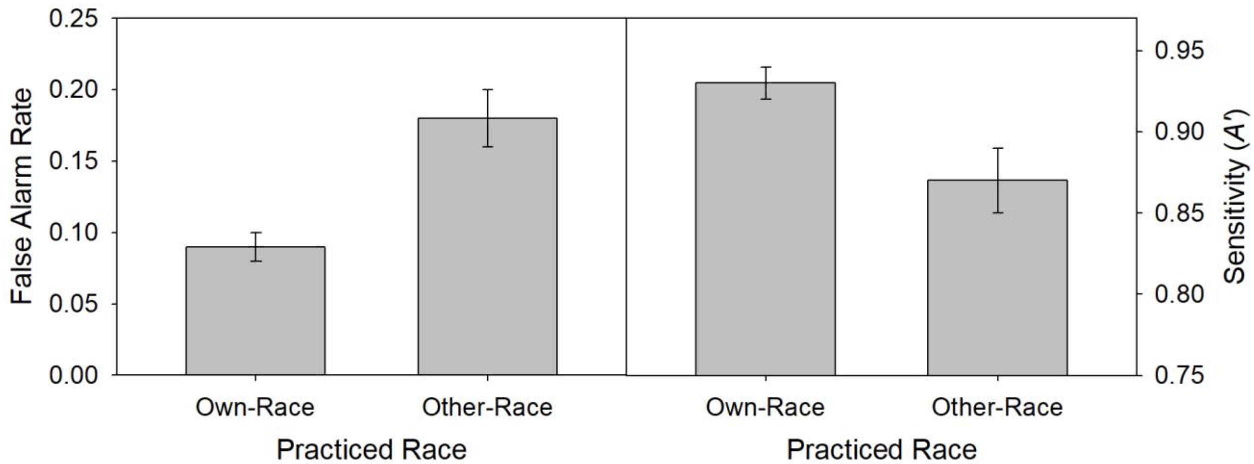

Figure 22. False alarm rates (left) and sensitivity (*A′*, right) across own- and other-race faces for participants who experienced retrieval practice.

Multidimensional Scaling Solutions

Using the SpAM method, participants' perceived similarity was assessed by inter-item distances in psychological space. First, average distance was assessed by separate 3 (Item Type) factor RM ANOVAs across Practice and Control conditions. I predicted that average inter-face distances would mirror predicted results from Experiment 2a: For Practice group participants, I predicted that average distances would differ as a function of the Item Type, with greater inter-face distances for Rp+ items, relative to baseline items, and smaller inter-face distances for Rp- items, relative to baseline items. For Practice group participants, there was no reliable effect of Item Type, $F(2, 104) < 1$, $p > .05$, $\eta^2 < .01$, with comparable distances across Nrp items (M = 508, SE = 9), Rp- items (M = 508, SE = 10), and Rp+ items (M = 509, SE = 10). To examine whether inter-face distances differed as a function of Practiced Race, average inter-face distances were also assessed by 3 (Item Type) x 2 (Practiced Race) mixed model ANOVAs, with Practiced

Race held between subjects. This analysis yielded no reliable effect of Item Type $F(2, 102) = .06$, $p > .05$, $\eta^2 < .01$,). In addition, there was no reliable interaction between Item Type and Practiced Race, $F(2, 102) = 1.63$, $p > .05$, $\eta^2 = .03$, illustrated in Figure 23. Although I predicted that average inter-face distances would mirror predicted results from Experiment 2a, Practice group participants did not show any differences as a function of RIF. This is consistent with the results of Experiment 1b, but inconsistent with the differences observed across Item Type from Experiment 2a.

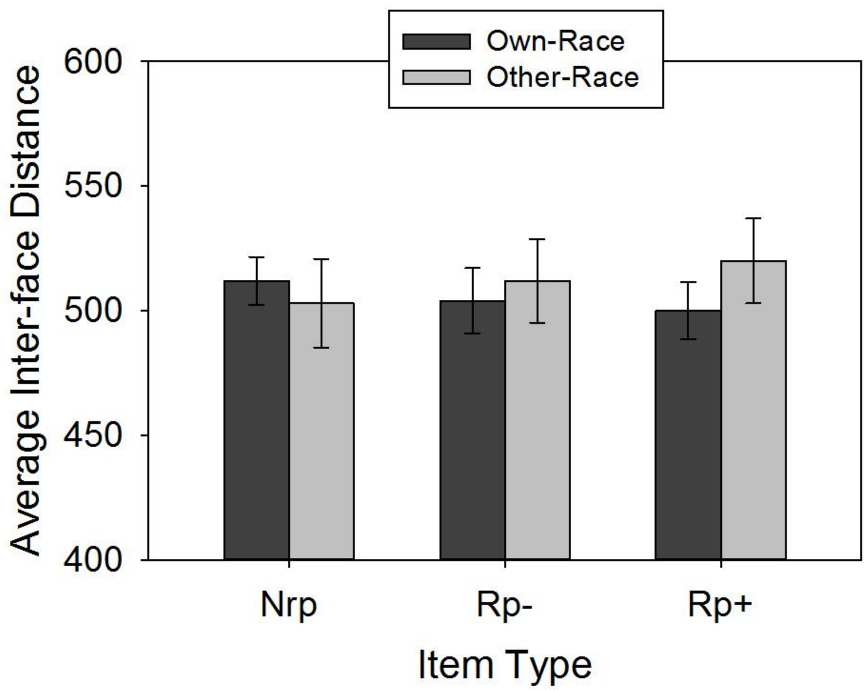

Figure 23. Average inter-face distance as a function of Practiced Race and Item Type for participants who experienced retrieval practice.

For Control group participants, I predicted that average distances would also mirror predicted results for Experiment 2a, with larger inter-face distances for restudied (S+) faces, but no differences across not restudied but related (S-) and nonstudied (Ns) faces. Average distance was assessed by a 3 (Item Type) factor RM ANOVA, and Control participants' average inter-face distances similarly revealed no reliable effect of Item Type, $F(2, 60) = .06$, $p > .05$, $\eta^2 = .02$.

To examine distance as a function of the Practiced Race, distances were also assessed through a 3 (Item Type) x 2 (Practiced Race) mixed model ANOVA, with Practiced Race held between subjects. Unlike Practice group participants, Control participants' inter-face distances showed reliable differences across Practiced Race, with participants who practiced Own-race faces showing reliably larger distances than participants who practiced Other-race faces, $F(1, 30) = 9.20$, $p < .05$, $\eta^2 = .24$, illustrated in Figure 24. Planned comparisons revealed these differences were only reliable between own-race S+ faces ($M = 542$, $SE = 14$) and other-race S+ faces ($M = 477$, $SE = 11$), $t(30) = 3.54$, $p < .05$, Cohen's $d = 1.26$, and between own-race S- faces ($M = 549$, $SE = 14$) and other-race S- faces ($M = 476$, $SE = 16$), $t(30) = 3.48$, $p < .05$, Cohen's $d = 1.23$.

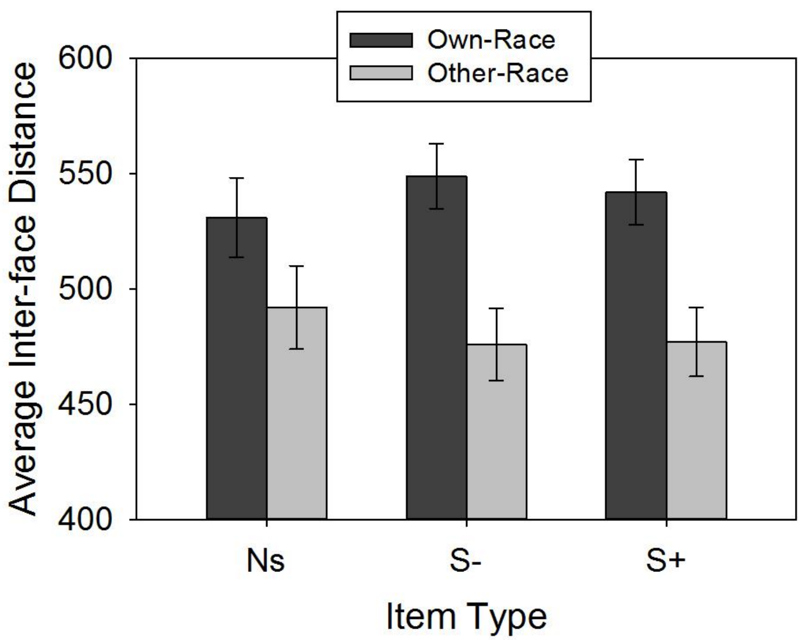

Figure 24. Average inter-face distance as a function of Restudied Race and Item Type for participants who experienced restudy conditions.

For participants who received no RIF component, average distances were again examined as a function of Race by a paired samples *t* test. This analysis revealed that although distances were numerically higher for Own-race identities, average distances were not reliably different across Own-race faces ($M = 469$, $SE = 18$) and Other-race faces ($M = 452$, $SE = 18$), $t(15) = 1.71$,

$p > .05$, Cohen's $d = .43$. To examine whether these distances differed as a function of RIF instructions, an additional 3 (Instruction: RIF, Control, MDS only) x 2 (Race: Own, other) mixed model ANOVA was conducted, with Instruction held between subjects. Like the comparison between Practice and Control conditions, this analysis yielded no reliable effects of Race, $F(1, 98) = 1.51$, $p > .05$, $\eta^2 = .01$. There was, however, a reliable effect of Instruction, $F(2, 98) = 3.88$, $p < .05$, $\eta^2 = .08$, and a reliable interaction, $F(2, 98) = 3.98$, $p < .05$, $\eta^2 = .08$. The results of this interaction revealed that average distances for participants who experienced MDS only instructions (i.e., no memory component) had overall smaller distances than participants in RIF or Control conditions. Overall, these results do not support the hypothesis that retrieval practice produces observable differences in psychological space, and, combined with the results of Experiment 1b, fail to provide support for SpAM as a sensitive index of RIF effects.

General Discussion

The purpose of the current experiments was to investigate the inhibitory explanation of retrieval-induced forgetting; in particular, these experiments examined the interference dependence assumption of the inhibitory account. Interference dependence describes the fact that baseline memory strength *should* influence the magnitude of RIF, with stronger memories producing greater competition during retrieval practice. To resolve this competition, inhibitory mechanisms must be activated to suppress representations of competing, but currently irrelevant, information, so that desired information can be retrieved successfully. The aim of Experiment 1 was to replicate and extend the recent finding that non-typical objects produce a larger magnitude RIF than typical objects (Reppa et al., 2017), potentially due to additional competition produced when non-typical objects are unpracticed competitors. Experiment 2 also aimed to replicate and extend recent findings showing that own- and other-race faces, which are more and less memorable, respectively, are also susceptible to RIF effects (Rugo et al., 2017). The results of Experiment 1 replicated three prior findings: RIF occurs for visual stimuli (Maxcey & Woodman, 2014; Reppa et al., 2017), these effects can be demonstrated using recognition tasks (Reppa et al., 2017; Rugo et al., 2017), and the magnitude of the RIF effect differs as a function of unpracticed, but related (i.e., Rp-), items' memory strength. Broadly, the results of Experiment 2 provide inconclusive evidence about the nature of face representations: Although participants showed some benefits of retrieval practice and restudy, they did not show the typical RIF effect in hit rates, due largely to no differences between hits for Rp- faces and Nrp faces.

The results of Experiment 1a demonstrated that, overall, participants showed a significant RIF effect across typical and non-typical items, with worse memory performance for Rp- items, and better memory for Rp+ items, relative to Nrp items. Critically, the RIF effect was larger for non-typical exemplars than it was for typical exemplars, suggesting that when non-typical items are not practiced or currently relevant, they nevertheless produce some competition during retrieval practice. Considered with evidence that strong competitors are more likely to show RIF than weak competitors (Anderson et al., 1994; Shivde & Anderson, 2001; Storm et al., 2007; see also Spitzer, 2014), and that non-typical items are more susceptible to RIF than typical items (Reppa et al., 2017), Experiment 1 supports the interference dependence prediction of the inhibitory account of RIF (but see Jakab & Raaijmakers, 2009; Williams & Zacks, 2001). These findings are also difficult to explain using associative accounts (Butler, Williams, Zacks, & Maki, 2001) and context-based accounts (Jonker et al., 2013; Jonker, Seli, & MacLeod, 2015) of RIF, given that these alternate views predict that items with strong category-exemplar associations (e.g., those between typical items and their category) should suffer more than weaker associations (e.g., those between non-typical items and their category).

The associate account of RIF suggests that when participants engage in retrieval practice, associations between practiced exemplars and category membership is strengthened. This strengthening of Rp+ items provides benefits during retrieval, but blocks access to Rp- items associated with the same category cue (Butler et al., 2001; Camp, Pecher, & Schmidt, 2007; Perfect et al., 2004). According to associative accounts, stronger links between category-exemplar pairs should result in larger magnitude RIF. By extension, this account would predict participants should show greater RIF for Typical Rp- items. On the other hand, context-based accounts of RIF argue that in order for RIF to occur, participants must experience a context change between study and retrieval practice, and that retrieval practice context must be reinstated during the final memory test (Jonker et al, 2013; Jonker et al., 2015). Although the context-based account makes no formal predictions about baseline item strength, memory for stronger category-exemplar pairs might be more disrupted by context changes (Sahakyan & Goodman, 2010; Jonker et al., 2013). The prediction that strong category cue-exemplar associations would be associated with greater RIF was not supported in the current experiment, as arguably *weaker* category-exemplar pairs (i.e., non-typical items) produced significant RIF, while stronger pairs (i.e., typical items) did not.

In addition, the results of Experiment 1a suggest that restudy of visual representations might produce similar RIF effects, albeit effects less sensitive to the influence of baseline item strength. Participants who simply restudied a subset of exemplars instead of experiencing retrieval practice also showed a significant difference in hit rates between items that were restudied (S+), related to restudied items but not restudied (S-), and unrelated items that were not restudied (Ns). Although these participants did not show differences in the magnitude of the effect based on typicality, this suggests that refreshing an item's episodic representation, as would be the case during restudy, might also produce some competition (e.g., through spreading activation; Collins & Loftus, 1975; Collins & Quillian, 1972). Relatedly, there is evidence that covert retrieval (e.g., participants considering whether a currently-viewed item has been seen before, even if they are not instructed to actively retrieve information) benefits memory in similar ways as overt retrieval (i.e., retrieval practice; Kang, 2010; Smith, 2011; Tulving, 1983; but see MacLeod, Gopie, Hourihan, Neary, & Ozkubo, 2010). Recent investigations of RIF effects have also shown that when participants engaged in mixed practice, such as when restudy and retrieval practice conditions are interleaved within the same practice block, RIF effects are observed (Dobler & Bäuml, 2013). If participants in the current study engaged in covert retrieval while restudying objects, it is possible that these retrieval processes would be sufficient to produce inhibition of Rp- items. If participants did engage covert retrieval, the fact that participants engaged in multiple restudy blocks that paralleled multiple retrieval practice blocks might have encouraged RIF effects for restudied items.

The evidence of difference between S+ and S- items in participants' hit rates also lends additional support to the inhibitory account of RIF, relative to context-based accounts of RIF (Jonker et al., 2013). According to context-based accounts, RIF requires two experimental elements: The study and practice phase must differ in context (e.g., passive study or pleasantness ratings followed by recognition practice), and the practice and test phases must have the same context (e.g., recognition). Context-based accounts argue that because Rp- items are associated with the practiced context, but were not actively practiced, the lack of context reinstatement during test impairs memory for Rp- items. Nrp items, which are only associated with the study context, do not benefit from context reinstatement at test, but they are also not impaired because they have only the study context. Rp+ items, which have their context reinstated across practice and test, then show benefits in memory performance. Recent outlines of the context account of RIF argue that restudy does not qualify as a context shift, and that is why no differences emerge between Ns and S- items at test (Jonker et al., 2015). However, control conditions of Experiment 1a are inconsistent with these assumptions: According to context-based accounts, participants did not experience a context shift between pleasantness ratings and restudy, yet memory performance still showed differences across S+, S-, and Ns exemplars. Considered with participants' *A′* scores, which did not show any difference across S+ and Ns exemplars, these findings provide mixed support for RIF of visual information when item representations are restudied (c.f. Anderson & Bell, 2001; Bäuml & Aslan, 2004; Ciranni & Shimamura, 1999; Staudigl et al., 2010), and do not provide support to context-based accounts of RIF (e.g., Jonker et al., 2013; Jonker et al., 2015).

The results of Experiment 2a, on the other hand, do not replicate recent findings that faces are susceptible to RIF (Ferreira et al., 2014; Rugo et al., 2017). Although there were numerical differences between hit rates for Nrp, Rp-, and Rp+ items, these differences did not reveal significant RIF for participants who experienced retrieval practice. In addition, participants who experienced restudy showed similar faciltation in hit rates for S+ faces, and no differences across Ns and S- faces. On the other hand, sensitivity (as *A'* scores) for participants who engaged retrieval practice did overall show RIF; however, this effect was driven by poor sensitivity for other-race Rp- faces. This suggests that other-race faces, which are associated with weaker memory representations, may be more susceptible to discrimination difficulties at test, rather than forgetting induced by retrieval practice. Considered with the fact that participants showed superior memory for Rp+ faces, regardless of the faces' race, this might instead suggest that lack of perceptual discriminability plays a role in producing RIF.

Although this does not replicate Rugo et al.'s (2017) finding that White faces (i.e., own-race) produce significant RIF, while Black (i.e., other-race) faces do not, these results are not entirely unprecedented. Because unfamiliar face memory is notoriously poor, relative to familiar face memory (Hill & Bruce, 1996; O'Toole et al., 1998; see also Burton, 2013; Johnston & Edmonds, 2009), these results imply competition between practiced, and unpracticed but related, exemplars may simply be unecessary for unfamiliar identities. When an unfamiliar face experiences benefits of retrieval practice or restudy, as evidenced by better memory for both Rp+ and S+ items, this process may not produce any competition between unpracticed exemplars because their memory traces are weak regardless of the identity's race. These findings also suggest that retrieval practice might not produce competition between studied exemplars because the amount of time spend refreshing the identity representation through practice is generally much shorter than time spent restudying the same information.

Although Experiment 2a failed to replicate the RIF effect on White faces observed by Rugo et al. (2017), the current study included several methodological departures. First, participants in the current study learned more identities than those used previously (20 faces of each race, instead of 10 faces of each race), and participants had less time to study each face (5 seconds instead of 10). It is possible that increasing the number of identities resulted in overall weaker memory traces for each studied face, and these weaker traces precluded competition during retrieval practice. Decreasing the study time from 10 seconds to 5 seconds is unlikely to have played a significant role in eliminating RIF effects, as eye-tracking evidence has demonstrated that participants produce more eye movements within few seconds of study and may taper off over time, not effectively utilizing extended study periods and particularly for other-race faces (see Goldinger, He, & Papesh, 2009). Additionally, Rugo et al. (2017) suggested that other-race identities, such as Asian faces studied by White participants, might be immune to RIF, because there are fewer other-race identities stored in memory, leading to sparser neural networks (Anderson, 1974). This in turn should lead to less competition during retrieval. The current experiment does not support this hypothesis, as only other-race (i.e., Asian) identities showed significant RIF in *A'* scores. Based on overall memory performance, the assumption that other-race faces are immune to RIF seems less likely than the explanation that face memory is generally poor, and might fail to produce inter-item competition during retrieval practice. In the future, to encourage participants to form stronger memory traces of studied identities, participants might show RIF effects if faces are learned across multiple study sessions, or if identities are used who are recognizable or highly memorable (e.g., celebrities, or unique identities with memorable characteristics).

The results of Experiment 1b and 2b do not provide conclusive evidence about the nature of item representations following RIF. Although participants' inter-object distances were smaller for typical, relative to non-typical items, there was no difference across Rp-, Nrp, and Rp+ items for participants who experienced retrieval practice, nor were there differences across S-, Ns, and S+ items for participants who experienced restudy. However, it is possible that because in Experiment 1a, participants showed similar patterns of RIF across Practice and Control conditions, no differences were present in Practice and Control MDS spaces. Similar to Experiment 1b, the results of Experiment 2b do not provide conclusive evidence about how episodic representations might change as a function of RIF. Participants did not show differences in average inter-face distances as a function of retrieval practice or restudy conditions. Although there was limited evidence that the Spatial Arrangement Method of MDS may be sensitive to other-race effects (see Meissner & Brigham, 2001) as a measure of psychological space (see also Papesh & Goldinger, 2010; Valentine, 1991), differences between own- and other-race faces were only present for participants who experienced restudy. Using SpAM, participants were instructed to drag and drop studied items and novel distractors in such a way that shorter distances reflected higher degrees of perceived similarity. Although SpAM produces MDS solutions that are comparable to those generated by traditional pairwise comparisons (Hout et al., 2013), there are several explanations as to why this method did not produce inter-object and inter-face distances that were sensitive to RIF effects.

First, because participants in Experiment 1b showed significant differences across inter-object distances based on typicality, it is possible that participants' psychological spaces were dominated by typicality (see, e.g., Figure 11), and that spatial arrangement was simply not sensitive enough to detect differences based on inhibitory mechanisms. Second, it is also possible that because SpAM involves participants having access to all exemplars while dragging and dropping into the active arena, item representations could be continually refreshed. Because RIF effects are temporary (MacLeod & Macrae, 2001; Storm, Bjork, & Bjork, 2011; see also Murayama et al., 2014), it is possible that this eliminated any effects of inhibition that might have been enacted during practice or restudy. Indirect measures that involve secondary measurements, such as same/different judgments about two exemplars, may provide a more sensitive method of investigating RIF effects in future investigations. Additionally, it is possible that inhibitory mechanisms do not act on episodic representations at all, resulting in a lack of differences across items beyond what was observed based on typicality. However, this explanation seems unlikely given that there is evidence of RIF acting upon episodic representation (Racsmány & Conway, 2006; Storm et al., 2007).

Finally, it is possible that RIF effects are simply not observable using non-explicit testing measures, as previous work has provided mixed evidence about the efficacy of using implicit measures in measuring RIF effects (Butler et al., 2001; Camp, Pecher, & Schmidt, 2005; Perfect et al., 2002). For example, Butler et al. (2001) and Perfect et al. (2002) failed to observe RIF effects when memory was assessed through fragment completion. However, Camp et al. (2005) provided evidence that implicit measures *can* reveal RIF effects, but only when participants are made explicitly aware of the relationship between experimental phases. Specifically, participants who reported they noticed the connection between study (category-exemplar pairs) and test (generating exemplars from categories that included those studied), but had not tried to use conscious memorization strategies, generated fewer Rp- items, relative to participants who reported they were unaware of any connection. Combined with the lack of effects in Experiment 1b and 2b, it is possible that participants must be made explicitly aware of the connection between study and MDS arrangement in order for differences in psychological space to be observed. To date, there have been no investigations of RIF of visual stimuli using implicit measures, so it is also possible that traditional implicit measures of verbal information (e.g., fragment completion, stem completion) are more likely to produce RIF than implicit measures that can be used to assess visual long-term memory (e.g., speeded perception tasks, MDS).

Considered together, the results of Experiment 1a and 2a provide support for the hypothesis that RIF affects an item's episodic representation (Anderson, 2003). This is also in line with neurological evidence suggesting that inhibitory mechanisms are activated during retrieval practice. For example, Johansson, Aslan, Bäuml, Gäbel, and Mecklinger (2007) investigated event-related potentials (ERPs) elicited during a RIF paradigm. Results demonstrated that ERPs elicited during retrieval practice over prefrontal regions (i.e., those involved in active inhibition; Shimamura, 2000; Anderson, 2003) were larger than those elicited during restudy. In addition, ERPs associated with retrieval practice were also predictive of RIF: Larger frontal amplitudes during retrieval practice were associated with greater forgetting. Additional evidence has shown that when participants engage in the final test after retrieval practice, neural correlates of access to stored representations are elicited (Bäuml et al., 2010; see also Wimber, Rutschmann, Greenle, & Bäuml, 2008). Rp- items are associated with greater activation in the left anterior ventrolateral prefrontal cortex (Wimber et al., 2008), which has been suggested to reflect the weakened state of Rp- item representations. Considered with evidence that the prefrontal cortex is implicated in episodic memory retrieval (see Cabeza, Ciaramelli, Olson, & Moscovitch, 2008, for a review), the current results are most consistent with an inhibition-based account of RIF. Similar to recent evidence showing that RIF effects can be obtained presenting specific visual stimuli (Ciranni & Shimamura, 1999; Maxcey, 2016; Maxcey & Woodman, 2014; Reppa et al., 2017; Rugo et al., 2017), and with evidence of episodic inhibition using verbal stimuli (Racsmány & Conway, 2006), the current findings suggest that RIF is sensitive to episodic representations (e.g., specific objects or identities).

The current experiments represent an attempt to examine RIF effects based on competitor strength. The attempt to provide an alternate quantification of how precisely item representations might change as a function of inhibition, however, remains to be established. The use of recognition practice as a means to produce RIF provides researchers with flexibility to incorporate a wide range of visual information, including stimuli such as objects, faces, and even specific feature representations (see Ferreira et al., 2014). Finally, while the current attempt to assess the representational differences across practiced, relative to unpracticed, exemplars using SpAM did not produce clear results, future research can focus on potentially more sensitive measures to detect these differences, to examine the influence of inhibitory mechanisms on visual information.

These results provide additional evidence that stronger memories may be more susceptible to forgetting when they are unpracticed competitors (Anderson et al., 1994; Storm et al., 2007; Reppa et al., 2017; Rugo et al., 2017; see also Spitzer, 2014). Specifically, these findings support the interference dependence prediction of the inhibitory account of RIF (see Storm & Levy, 2012), and that RIF effects can be produced for visual information via recognition practice (Maxcey, 2016; Maxcey & Woodman, 2014; Reppa et al., 2017; Rugo et al., 2017). The inhibitory account of RIF proposes four primary predictions: Cue independence, retrieval dependence, strength independence, and interference dependence. Critical to the current study was interference dependence, or the assumption that the presence or magnitude of RIF is dependent on the amount of competition produced between category exemplars during retrieval practice. Items that create greater degrees of competition, whether through strong category-cue associations (Anderson et al., 1994) or through baseline item strength (Reppa et al., 2017; Rugo et al., 2017; Spitzer, 2014), enact greater need for inhibition. In essence, the strength or memorability of Rp- items is a better predictor of RIF than the degree to which Rp+ items are strengthened through practice. Competing perspectives of RIF, such as associative-blocking accounts (Butler et al., 2001; Camp et al., 2007; Perfect et al., 2004) or context-based accounts (Jonker et al., 2013), predict instead that association or similarity between category-exemplar pairs produce costs for Rp- items. Although RIF for memorable stimuli is not entirely incompatible with context-based accounts, the observed RIF for non-typical objects in both retrieval practice and restudy conditions are best explained by interference dependence. In addition, the finding that unfamiliar faces, regardless of race, did not show RIF supports inhibition-based accounts. Context-based accounts might argue that reinstatement of practice context should have produced impaired memory for Rp- faces regardless of their memorability. Lack of RIF for unfamiliar faces suggests instead that if there is no need for inhibitory processes to be enacted due to weaker memory traces, no deficit for Rp- items is observed. In sum, similar to recent evidence showing RIF effects are influenced by competitor strength (Reppa et al., 2017; Storm et al., 2007; see Spitzer, 2014 for a review), the current results suggest that stronger memories may in fact be more susceptible to forgetting.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451-474.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49,* 415–445.

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In C. MacLeod & B. Uttl (Eds.), *Dynamic Cognitive Processes* (pp. 159–190). Tokyo, Japan: Springer.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: the role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General*, *130*(3), 544.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic bulletin & review, 7*(3), 522-530.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: memory retrieval as a model case. *Psychological Review*, *102*(1), 68-100.

Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *The Quarterly Journal of Experimental Psychology*, *61*(5), 690-707.

Balas, B. (2012). Bayesian face recognition and perceptual narrowing in face-space. *Developmental Science*, *15*(4), 579-588.

Barnier, A., Hung, L., & Conway, M. (2004). Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition and Emotion*, *18*(4), 457-477.

Bäuml, K. H. (2002). Semantic generation can cause episodic forgetting. *Psychological Science*, *13*(4), 356-360.

Bäuml, K. H., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition*, *32*(4), 610-617.

Bäuml, K. H. T., & Samenieh, A. (2010). The two faces of memory retrieval. *Psychological Science, 21*(6), 793-795.

Bäuml, K.-H., & Samenieh, A. (2012). Selective memory retrieval can impair and improve retrieval of other memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 488–494.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In Robert L. Solso (ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123-144). Mahwah, NJ: Lawrence Erlbaum.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. *Varieties of Memory & Consciousness*, 309-330.

Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1998). Varieties of goal-directed forgetting. *Intentional forgetting: Interdisciplinary approaches* (*pp.* 103-137). Mahwah, NJ: Erlbaum.

Bjork, E. L., Bjork, R. A., & MacLeod, M. D. (2006). Types and consequences of forgetting: Intended and unintended. *Memory and Society: Psychological Perspectives*, 134-158.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325-14329.

Brown, J. (1976). An analysis of recognition and recall and of problems in their comparison. In J. Brown (Ed.), *Recall and Recognition.* Oxford, England: John Wiley & Sons.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243-248.

Butler, K. M., Williams, C. C., Zacks, R. T., & Maki, R. H. (2001). A limit on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1314-1319.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience*, *9*(8), 613-625.

Camp, G., Pecher, D., & Schmidt, H. G. (2005). Retrieval-induced forgetting in implicit memory tests: The role of test awareness. *Psychonomic Bulletin & Review*, *12*(3), 490-494.

Camp, G., Pecher, D., & Schmidt, H. G. (2007). No retrievalinduced forgetting using item-specific independent cues: Evidence against a general inhibitory account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 950–958

Carroll, M., Campbell-Ratcliffe, J., Murnane, H., & Perfect, T. (2007). Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology*, *19*(4-5), 580-606.

Catz, O., Kampf, M., Nachson, I., & Babkoff, H. (2009). From theory to implementation: Building a multidimensional space for face recognition. *Acta Psychologica*, *131*(2), 143-152.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1403-1414.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407-428.

Collins, A. M., & Quillian, M. R. (1972). Experiments on semantic memory and language comprehension. In L. W. Gregg, *Cognition in learning and memory.* Oxford, England: John Wiley & Sons.

Coman, A., & Hirst, W. (2015). Social identity and socially shared retrieval-induced forgetting: The effects of group membership. *Journal of Experimental Psychology: General*, *144*(4), 717.

Dobler, I. M., & Bäuml, K. H. T. (2013). Retrieval-induced forgetting: Dynamic effects between retrieval and restudy trials when practice is mixed. *Memory & Cognition*, *41*(4), 547-557.

Fan, J. E., & Turk-Browne, N. B. (2013). Internal attention to features in visual short-term memory guides object learning. *Cognition*, *129*(2), 292-308.

Faerber, S. J., Kaufmann, J. M., Leder, H., Martin, E. M., & Schweinberger, S. R. (2016). The role of familiarity for representations in norm-based face space. *PloS one*, *11*(5), e0155380.

Faye, P., Brémaud, D., Daubin, M. D., Courcoux, P., Giboreau, A., & Nicod, H. (2004). Perceptive free sorting and verbalization tasks with naive subjects: An alternative to descriptive mappings. *Food Quality and Preference*, *15*(7-8), 781-791.

Faye, P., Brémaud, D., Teillet, E., Courcoux, P., Giboreau, A., & Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference*, *17*(7-8), 604-614.

Ferreira, C. S., Marful, A., & Bajo, T. (2014). Interference resolution in face perception and name retrieval. *Acta psychologica*, *153*, 120-128.

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: insights from eye movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1105-1122.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178-200.

Gómez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition*, *33*(8), 1431-1441.

Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 986.

Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the Massive Memory picture database. *PLoS One*, *9*(11), e112644.

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, *142*(1), 256-281.

Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(1), 93-103.

Hugenberg, K., Young, S. G., Bernstein, M. J., & Sacco, D. F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological review*, *117*(4), 1168.

Jaworska, N. (2015). Chupetlovska-Anastova 2009: A Review of Multidimensional Scaling (MDS) and its Utility in Various Psychological Domains. *Tutorials in Quantitative Methods for Psychology. University of Ottawa. Viitattu*, *3*.

Johansson, M., Aslan, A., Bäuml, K. H., Gäbel, A., & Mecklinger, A. (2006). When remembering causes forgetting: Electrophysiological correlates of retrieval-induced forgetting. *Cerebral Cortex*, *17*(6), 1335-1341.

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological Science, 15*, 448–453.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*(5), 577-596.

Johnston, R. A., Milne, A. B., Williams, C., & Hosie, J. (1997). Do distinctive faces come from outer space? An investigation of the status of a multidimensional face-space. *Visual Cognition*, *4*(1), 59-67.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243-275.

Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting in context: an inhibition-free, context-based account. *Psychological Review*, *120*(4), 852.

Jonker, T. R., Seli, P., & MacLeod, C. M. (2015). Retrieval-induced forgetting and context. *Current Directions in Psychological Science*, *24*(4), 273-278.

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. Memory & Cognition, 38, 1009-1017.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1-27.

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification:

Testing the face-space framework. *Psychological Science*, *11*(5), 379-385.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in cognitive sciences*, *6*(7), 299-305.

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science*, *18*(1), 29-34.

LSU Fall Facts. (2018). Retrieved from https://www.lsu.edu/bgtplan/facts/pdfs/2018fallfacts_final.pdf.

Louisiana Population. (2019). Retrieved from http://worldpopulationreview.com/states/louisiana-population/.

Ma, Correll, & Wittenbrink (2015). The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. *Behavioral Research Methods*, *47*, 1122-1135.

Macleod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 671-685.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science*, *12*(2), 148-152.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology*, *77*(3), 463.

Maxcey, A. M. (2016). Recognition-induced forgetting is not due to category-based set size. *Attention, Perception, & Psychophysics*, *78*(1), 187-197.

Maxcey, A. M., & Bostic, J. (2015). Activating learned exemplars in children impairs memory for related exemplars in visual long-term memory. *Visual Cognition*, *23*(5), 643-658.

Maxcey, A. M., Bostic, J., & Maldonado, T. (2016). Recognition practice results in a generalizable skill in older adults: Decreased intrusion errors to novel objects belonging to practiced categories. *Applied Cognitive Psychology*, *30*(4), 643-649.

Maxcey, A. M., Glenn, H., & Stansberry, E. (2017). Recognition-induced forgetting does not occur for temporally grouped objects unless they are semantically related. *Psychonomic Bulletin & Review*, 1-17.

Maxcey, A. M., & Woodman, G. F. (2014). Forgetting induced by recognition of visual images. *Visual cognition*, *22*(6), 789-808.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3-35.

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, *1405*(5), 1383-1409.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 70-84.

Nørby, S. (2015). Why forget? On the adaptive value of memory loss. *Perspectives on Psychological Science*, *10*(5), 551-578.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., McKee, K., Huff, D., & Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. *Memory & Cognition*, *26*(1), 146-160.

Papesh, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own-and cross-race face spaces. *Cognition*, *116*(2), 283-288.

Perfect, T. J., Moulin, C. J., Conway, M. A., & Perry, E. (2002). Assessing the inhibitory account of retrieval-induced forgetting with implicit-memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1111-1119.

Perfect, T. J., Stark, L. J., Tree, J. J., Moulin, C. J., Ahmed, L., & Hutter, R. (2004). Transfer appropriate forgetting: The cue-dependent nature of retrieval-induced forgetting. *Journal of Memory and Language*, *51*(3), 399-417.

Raaijmakers, J. G., & Jakab, E. (2013). Rethinking inhibition theory: On the problematic status of the inhibition theory for forgetting. *Journal of Memory and Language*, *68*(2), 98-122.

Racsmány, M., & Conway, M. A. (2006). Episodic inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 44-57.

Reppa, I., Williams, K. E., Worth, E. R., Greville, W. J., & Saunders, J. (2017). Memorable objects are more susceptible to forgetting: Evidence for the inhibitory account of retrieval-induced forgetting. *Acta Psychologica*, *181*, 51-61.

Roediger, H. L., III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior, 12*, 44–657.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382-439.

Rugo, K. F., Tamler, K. N., Woodman, G. F., & Maxcey, A. M. (2017). Recognition-induced forgetting of faces in visual long-term memory. *Attention, Perception, & Psychophysics*, *79*(7), 1878-1885.

Saunders, J., Fernandes, M., & Kosnes, L. (2009). Retrieval-induced forgetting and mental imagery. *Memory & Cognition*, *37*(6), 819-828.

Sahakyan, L., & Goodmon, L. B. (2010). Theoretical implications of extralist probes for directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 920-937.

Sharman, S. J. (2011). Retrieval-induced forgetting of performed and observed bizarre and familiar actions. *Experimental Psychology*, *58*(5), 361-369.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review*, 2, 249–253.

Shepard, R. N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*(2), 125-140.

Shivde, G., & Anderson, M. C. (2001). The role of inhibition in meaning selection: Insights from retrieval-induced forgetting. *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity*, *1*, 175-190.

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology,* *28*(2), 207-218.

Smith, M., "Covert Retrieval Practice Benefits Retention As Much As Overt Retrieval Practice" (2011). All Theses and Dissertations (ETDs). 489.

Spitzer, B. (2014). Finding retrieval-induced forgetting in recognition tests: a case for baseline memory strength. *Frontiers in Psychology*, *5*, 1102.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology*, *25*(2), 207-222.

Staudigl, T., Hanslmayr, S., & Bäuml, K. H. T. (2010). Theta oscillations reflect the dynamics of interference in episodic memory retrieval. *Journal of Neuroscience*, *30*(34), 11356-11362.

Storm, B. C. (2011). Retrieval-induced forgetting and the resolution of competition. In *Successful Remembering and Successful Forgetting* (pp. 107-124). Psychology Press.

Storm, B. C., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? *Psychonomic Bulletin & Review*, *13*(6), 1023-1027.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2007). When intended remembering leads to unintended forgetting. *The Quarterly Journal of Experimental Psychology*, *60*(7), 909-915.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2012). On the durability of retrieval-induced forgetting. *Journal of Cognitive Psychology*, *24*(5), 617-629.

Storm, B. C., & Jobe, T. A. (2012). Remembering the past and imagining the future: Examining the consequences of mental time travel on memory. *Memory*, *20*(3), 224-235.

Storm, B. C., & Nestojko, J. F. (2010). Successful inhibition, unsuccessful retrieval: Manipulating time and success during retrieval practice. *Memory*, *18*(2), 99-114.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457-482.

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University Press.

Tulving, E., & Arbuckle, T. Y. (1963). Input and output interference in short-term associative memory. *Journal of Verbal Learning and Verbal Behavior, 1*, 321–334.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161-204.

Valentine, T. (2001). Face-space models of face recognition. *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges*, 83-113.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, *69*(10), 1996-2019.

Veling, H., & van Knippenberg, A. (2004). Remembering can cause inhibition: retrieval-induced inhibition as cue independent process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 315-318.

Waldhauser, G. T., Johansson, M., & Hanslmayr, S. (2012). Alpha/beta oscillations indicate inhibition of interfering visual memories. *Journal of Neuroscience*, *32*(6), 1953-1961.

Wimber, M., Bäuml, K. H., Bergström, Z., Markopoulos, G., Heinze, H. J., & Richardson-Klavehn, A. (2008). Neural markers of inhibition in human memory retrieval. *Journal of Neuroscience*, *28*(50), 13419-13427.

Wish, M., & Carroll, J. D. (1974). Applications of individual differences scaling to studies of human perception and judgment. *Handbook of Perception*, *2*, 449-491.

Appendix A. Massive Memory MDS Objects

Contained in Table A1 is a list of target categories and filler categories from the Massive Memory MDS Database (Hout et al., 2014). These categories were chosen to represent images that are distinctive based on shape, as object stimuli appeared in greyscale. Sample stimuli from the categories birdhouses, benches, and televisions can be found in Figure 2.

Table A1

Target Categories	Filler Categories
Birdhouses	Desks
Coffee makers	Printers
Ice skates	Dollhouses
Backpacks	Door knockers
Benches	Handbags
Bottles	Jackets
Cameras	Locks
Clocks	Phones
Dumbbells	Trophies
Fans	
Garbage cans	
Lamps	
Lawn mowers	
Speakers	
Televisions	
Vases	

Appendix B. Institutional Review Board Approval

ACTION ON EXEMPTION APPROVAL REQUEST

LSU

Institutional Review Board
Dr. Dennis Landin, Chair
130 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu
lsu.edu/research

TO: Megan Papesh
Psychology

FROM: Dennis Landin
Chair, Institutional Review Board

DATE: January 18, 2019

RE: **IRB# E11447**

TITLE: Retrieval-Induced Forgetting of Objects and Faces

New Protocol/Modification/Continuation: New Protocol

Review Date: 1/18/2019

Approved X **Disapproved** __________

Approval Date: 1/18/2019 **Approval Expiration Date:** 1/17/2022

Exemption Category/Paragraph: 2b

Signed Consent Waived?: No

Re-review frequency: (three years unless otherwise stated)

LSU Proposal Number (if applicable):

By: Dennis Landin, Chairman ____________________

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –
Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. SPECIAL NOTE: **When emailing more than one recipient, make sure you use bcc. Approvals will automatically be closed by the IRB on the expiration date unless the PI requests a continuation.**

* *All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb*

Appendix C. Approved Informed Consent

Consent Form

1. Study Title: Retrieval-Induced Forgetting of Objects and Faces

2. Performance Site: Louisiana State University

3. Investigators: Dr. Megan Papesh (mpapesh@lsu.edu) and Laura Heisick (lheisi1@lsu.edu).

4. Purpose of the Study: The purpose of this research project is to examine human memory, including encoding through to retrieval, and how successful retrieval may have unintended consequences on later memory.

5. Subject Inclusion: Native English-speaking individuals between the ages of 18 and 65 with normal or corrected-to-normal vision and normal color vision.

6. Number of subjects: 400

7. Study Procedures: The study will take between 30 and 60 minutes. Participants in this study will view a series of everyday objects and/or faces. Some participants will rate items based on their perceived pleasantness. Participants may see some items repeatedly. Participants may be asked to make subsequent decisions about previously viewed items, including comparisons of how similar they are.

8. Benefits: The data from this study will further research on human memory.

9. Risks: There are no known risks from participating in the current study.

10. Right to Refuse: Subjects may choose not to participate or to withdraw from the study at any time without penalty or loss of any benefit to which they might otherwise be entitled.

11. Privacy: Results of the study may be published, but no names or identifying information will be included in the publication. Subject identity will remain confidential unless disclosure is required by law.

12. Financial information: You will receive no financial compensation for participating in this study. You will receive one experimental credit for each half hour of participation.

13. Withdrawal: Your participation is voluntary, and you may withdraw from the study at any time and for any reason. If you decide not to participate or if you withdraw from the study, there is no penalty or loss of benefits to which you are otherwise entitled.

14. Removal: You can be removed from the experiment for behavior that is disruptive or harmful to others.

15. Signatures: The study has been discussed with me and all my questions have been answered. I may direct additional questions regarding study specifics to the investigators. If I have questions about subjects' rights or other concerns, I can contact Dennis Landin, Institutional Review Board, (225) 578-8692. I agree to participate in the study described above and acknowledge the investigator's obligation to provide me with a signed copy of this consent form.

Signature of Subject ______________________ Date ______________

Vita

Laura Heisick received her bachelor's degree in Psychology from Louisiana State University in 2014. She was subsequently accepted into the LSU Cognitive and Brain Sciences doctoral program, where she completed her M.A. in 2017. Her research interests broadly include memory, including successful retrieval and memory failures such as forgetting, as well as face perception, with emphasis on the difficulty observers have when viewing unfamiliar identities. She will receive her Ph.D. in August 2019 and plans to pursue a position in academia upon graduation.