Context reinstatement reconsidered: Investigating boundary conditions of the effect

by

Christopher Michael Lee

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner

Dr. Jennifer Ryan
Senior Scientist
Rotman Research Institute and Canada Research Chair in Cognitive Neuroscience of Memory
Professor
Departments of Psychology and Psychiatry
University of Toronto

Supervisor

Dr. Myra A. Fernandes
Professor
Department of Psychology
University of Waterloo

Internal Member(s)

Dr. Colin M. MacLeod
Professor
Department of Psychology
University of Waterloo

Dr. Roxane J. Itier
Associate Professor
Department of Psychology
University of Waterloo

Internal-external Examiner

Dr. Ewa Niechwiej-Szwedo
Associate Professor
Department of Kinesiology
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Christopher Lee was the sole author for Chapters 1, 2, 3 and 5 which were written under the supervision of Dr. Myra Fernandes and were not written for publication.

This thesis also consists in part of one manuscript written for publication. At the time of the submission of this thesis, the manuscript has been submitted for publication and is under review. Exceptions to sole authorship of material are as follows:

Research presented in Chapter 4:

This research was conducted at the University of Waterloo by Christopher Lee and Ryan Yeung, under the supervision of Dr. Myra Fernandes. Both Christopher Lee and Ryan Yeung contributed to the study design, procedure, data analysis and manuscript drafts.

Abstract

The context reinstatement (CR) effect suggests that target items are easier to recognize when encoding and retrieval contexts are matched. In this PhD thesis, I manipulated features of the target and the context, in target-context pairs, to examine the boundary conditions of the CR effect. In Experiments 1-3, I investigated whether familiarity with the target item would influence the CR magnitude for target words (Experiment 1), objects (Experiment 2) and faces (Experiment 3). Familiarity was manipulated using a pre-exposure phase during which participants viewed targets either 0, 1, 3 or 10 times. During a later encoding phase, targets were paired with a unique indoor or outdoor scene and participants indicated how likely the target and context matched. During a retrieval phase, targets and lures were presented during each trial of a recognition test, paired with either the same, or with a new context. A significant CR benefit was observed for target words (Experiment 1) and objects (Experiment 2), and the pre-exposure manipulation did not influence the magnitude of the memory benefit, in line with 'global matching' models of memory. Importantly, I observed a significant reduction in the CR benefit for faces (Experiment 3) that were highly familiar, that is pre-exposed 10 times, compared to 0 times. In Experiment 4, I reduced the number of pre-exposures in the most familiar condition, from 10 to 5, to rule out the possibility that a ceiling effect, for face memory performance, could account for this finding. Again, I observed a reduction in the CR benefit for faces that were highly familiar, in line with an 'outshining' model of memory.

In Experiment 5, I investigated an attentional account for this stimulus specific effect on the CR benefit to memory. Participants completed a paradigm similar to that used in Experiments 1-4, but memory was assessed for target words and faces in separate blocks in a within-subjects design. Prior to each block, eye gaze was calibrated with an eye tracker as an

indirect measure of overt attention. Replicating my previous findings, the CR benefit for target faces gradually diminished as familiarity (number of pre-exposures) with the face increased. Importantly, eye tracking data gathered during the encoding phase revealed that for face-context pairs, a higher proportion of fixations and dwell time was placed on the face than the context. The opposite pattern was revealed for word-context pairs. Results suggest that devoting more attention to the context than the target, during encoding, may engender a necessity for context to inform memory decisions, accounting for the consistent benefit of CR on memory for target words, but not target faces. Faces, in contrast, are not as reliant on context to guide memory performance.

In Experiments 6 and 7, I investigated whether the anxiety-provoking nature of a context scene might influence the CR benefit to memory. During encoding, participants viewed target faces paired with scenes validated as either highly anxiety-provoking or not, half of which contained other faces embedded within the scene. At retrieval, target faces were presented again with either the same or a new context scene. The expected CR benefit was observed when the contexts were low-anxiety scenes or high-anxiety scenes without embedded faces. In contrast, the CR benefit was significantly reduced when the contexts were high-anxiety scenes containing embedded faces. Results suggest that the benefit of reinstating a context, on memory for target faces, depends critically on semantic characteristics of the reinstated context.

My PhD research suggests a refinement to the classic CR effect: it applies primarily to memory for words or objects and for faces that are relatively novel. When faces are familiar, or encoded in highly anxiety-provoking scenes, other factors, such as target signal strength are more important in guiding recognition memory decisions.

Acknowledgements

I would first and foremost like to thank my supervisor and mentor, Dr. Myra Fernandes. She has been patient and helpful, not only in my professional career, but also in understanding all of the hardships I have endured these past few years. Whether it be a severely injured back, or a broken computer, Dr. Fernandes has always been understanding and supportive and willing to help me get through. Professionally, I never would have made it to this point without all of her guidance. Her wise advice and sage criticisms have been instrumental in moulding me into the researcher I am today.

I would also like to thank the rest of my thesis committee, Dr. Colin MacLeod, Dr. Roxane Itier, Dr. Jennifer Ryan and Dr. Ewa Niechwiej-Szwedo, for all of their insightful comments and challenging discussions.

I would next like to thank my family. Their endless, unconditional support is what motivated me to pursue these studies in the first place. They have always pushed me to better myself, to work hard and to not be afraid to be whatever I wish to be.

Last, but most definitely not least, I would like to thank my partner, Kingsley. She has been an integral part of my life and has made me a better person. She has always made me strive for the best and to never settle for less. Even when I was up late at night working on this thesis, she was always supportive and there for me whenever I was stressed or just needed a break. I could not have done this without you!

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Imagine this scenario. You are walking through the grocery store, picking up food for your evening meal. As you carry on, you notice an individual from the corner of your eye. You immediately take notice and register them as somewhat familiar. The difficulty, however, is you cannot seem to quite place them. At least, not until you return home only to realize that they are the barista at the local coffee shop you frequent. You may feel upset that you were unable to recognize this individual that you interact with on a near daily basis. Fear not, for this phenomenon has a firm basis in the Psychological memory literature.

The context reinstatement (CR) effect describes the exact phenomenon demonstrated in the scenario above. This effect proposes that memory for an individual, or any to-be-remembered item, is enhanced when the individual is re-encountered within the exact same context as it was initially learned or encoded in. That is, information is easier to remember when the encoding and retrieval contexts are matched (Dalton, 1993; Dougal & Rotello, 1999; Godden & Baddeley, 1975; Gruppuso, Lindsay, & Masson, 2007; Hockley, 2008; Macken, 2002; Mandler, 1980; Murnane & Phelps, 1993; Smith, 1988, 1994; Smith & Vela, 2001). Possibly the best demonstration of this effect is the well known butcher-on-the-bus scenario (Mandler, 1980). In this scenario, one’s recognition memory for a butcher can be enhanced, and is richer (accompanied by their name and details from past encounters) when encountered in the context of their butcher shop. However, if seen on a bus (i.e. in a different context), the butcher is more difficult to recognize. The argument is that recognition memory is accompanied only by a sense of familiarity and an absence of contextual details such as the person’s identity. But how reliable is the CR effect, and what factors influence the magnitude of the memory benefit that context confers?

1.1 What is the CR effect?

To date, a large number of studies and theoretical accounts have contributed to the literature on the CR effect. Context effects have been observed using a variety of target stimuli, though much of the early research examines its effect on memory for words and word pairings (Emmerson, 1986; Fernandez & Glenberg, 1985; Humphreys, Pike, Bain, & Tehan, 1989; Murnane & Phelps, 1993, 1994, 1995; Murnane, Phelps, & Malmberg, 1999; Smith, 1979, 1985, 1986; Smith, Glenberg, & Bjork, 1978; see Smith & Vela, 2001 for review). In general, words presented at test in the same or similar context as the study episode are more likely to be remembered on a recognition or recall test, than are words presented in a novel context (Godden & Baddeley, 1980; Rutherford, 2004). In the seminal study conducted by Godden and Baddeley (1975), participants studied a list of words either on land or underwater. They were then tested either in the same or the other context. Participants demonstrated improved memory performance when the encoding and test contexts matched. In line with the butcher-on-the-bus scenario, word recall in a context different to that of the study episode reduced the contribution of recollection (memory for details from encoding) to memory judgments (Macken, 2002).

In more recent work, this effect of context on recognition memory has been demonstrated using other types of stimuli, and in multiple populations. Craik and Schloerscheidt (2011) presented object names or object pictures on background scenes (e.g., outdoor landscapes) to younger and older adults. Participants later attempted to recognize previously presented items on background scenes that were the same, switched, new, or blank. Their findings indicated that performance was better in the same relative to the new context condition (CR effect), showing that CR effects are reliable across different material types, and age groups.

Research has also found memory for faces to be influenced by context. Gruppuso and colleagues (2007) presented participants with images of faces paired at encoding with unique context scenes, including buildings and scenery. Participants were then given a surprise recognition memory test for the studied faces, paired during retrieval with either the *same* context as at encoding, a *switched* context (a contest previously seen with a different face), or a *new* context (a novel context not from the study phase); new faces on the recognition test were paired with either a context previously seen (but with a different face) or a *new* context. Results showed a significant boost in subjective, recollection-based, memory for faces paired with the same relative to a new context at test—the well-known CR effect, highlighting that CR effects extend to memory for faces.

1.2 Role of Familiarity and Inconsistencies in Reports of the CR Effect

There is evidence that certain factors, such as meaningfulness of the target or encoding time, can influence the magnitude of benefit conferred by context by affecting the familiarity of the target. For example, Isarida, Isarida, and Sakai (2012) investigated whether target words and non-words, encoded for either short or long durations, would receive a similar memory benefit from context reinstatement. In their experiments, CR effects were observed for words that had been encoded for a short (1.5 s), but not for a long (4 s) duration. They suggested that the increase in encoding time rendered the target words more meaningful, to the point that the context was no longer needed to enhance memory. In contrast, they showed that non-words, compared to real words, received a reliable benefit from context reinstatement, regardless of encoding duration. They argued that non-words are unfamiliar and thus always benefit from reinstatement of context. They concluded that CR effects decrease with increasing familiarity, here defined as meaningfulness of the target materials. Their results are discussed as support for an “outshining account” of memory. That is, as target items become more familiar, context becomes less beneficial in further enhancing memory (Smith & Vela, 2001). The familiar targets “outshine” the contexts. In support, Schonbach-Medina and Vakil (2017) also found that extending the exposure time of a set of words, which they argued increases familiarity, correspondingly reduced context effects on memory.

Converging research has also shown that recognition of unfamiliar or meaningless materials, such as unfamiliar faces or non-words, benefits reliably from reinstatement of context at retrieval (e.g., Dalton, 1993; Malpass & Devine, 1981; Russo, Ward, Geurts, & Scheres, 1999). For example, using faces as targets, Dalton (1993) showed a significant CR benefit on recognition of unfamiliar faces. This effect was later conceptually replicated by Russo and colleagues (1999), who also found similar CR effects on recognition of non-words (but not with meaningful words). In line with this, eyewitness research has shown CR benefits in recognition of unfamiliar faces of culprits of crimes (e.g., Krafka & Penrod, 1985; Malpass & Devine, 1981; Smith & Vela, 1992). These studies provide evidence that familiarity, manipulated by varying meaningfulness or encoding time, may be an important factor influencing the magnitude of CR effects.

There is, however, also evidence to suggest that the role of familiarity can vary depending on the way in which it is manipulated. Murnane and Phelps (1995) demonstrated, in a series of studies, that increasing the cue strength (familiarity) of a target word via repetition and deeper processing (Craik & Lockhart, 1972) increased the magnitude of the CR effect. As well, Reder and colleagues (2013) have suggested that highly familiar (famous) targets are more easily bound to contexts because they have pre-existing representations, and that this enhances CR effects. Their experiments showed that the CR effect is particularly beneficial for faces with pre-existing memory representations (i.e. famous faces), relative to novel/unfamiliar ones. These findings are contrary to the pattern shown by other studies (e.g. Isarida et al., 2012; Schonbach-Medina & Vakil, 2017).

It must be noted that, in addition to inconsistent methods of familiarity manipulation, some of the studies just outlined used words as their target stimuli (Isarida et al., 2012; Murnane & Phelps, 1995; Russo et al., 1999; Schonbach-Medina & Vakil, 2017) whereas others used faces (Dalton, 1993; Reder et al., 2013). The studies above also manipulate 'context' in varying ways. Some use different rooms or experimenters (Dalton, 1993; Isarida et al., 2012); some odours (Dalton, 1993); some word pairs (Schonbach-Medina & Vakil, 2017); some visual scenes (Reder et al., 2013); and still others use a combination of foreground colour, background colour and screen location (Murnane & Phelps, 1993; 1994; 1995) as context. Therefore, not only are there inconsistencies in findings based on the way that memory is tested (Godden & Baddeley, 1975; 1980), but key factors, including the target, the context and familiarity manipulations, are all inconsistent across experiments. With such a large variety in nearly every aspect of experiments contributing to the context reinstatement literature, it is difficult to truly compare experiments to one another and to identify critical factors influencing the benefit of context effects.

To advance our understanding of the CR effect, it is important to not only consider what its capabilities are, but also any limits or *boundary conditions* on the phenomenon. Familiarity, as discussed above, does seem to pose a limitation on the benefit context can provide. However, to this point, studies have mainly assessed the role of familiarity using a dichotomy or two separate familiarity categories (i.e. famous and non-famous faces). The goal of Chapters 2 and 3 of this dissertation is to use a quantifiable means of manipulating degree of familiarity that a participant has with a target item. Additionally, no research to this point has directly compared how familiarity with various stimulus types can uniquely impact the CR effect. In Chapter 2, I investigate how familiarity with words, objects and faces, can influence the CR benefit. Next, in Chapter 3, I examine whether an attentional account may offer an explanation for the patterns of memory results observed in Chapter 2. In Chapter 4, I investigated how characteristics of the context itself could also influence the magnitude of CR benefit. In that chapter, the degree to which the contexts are social anxiety-provoking was manipulated, along with assessments of individual differences in trait social anxiety, to determine whether these influence the benefit of context reinstatement on target memory.

1.3 Theoretical Accounts

There are a number of theoretical explanations that attempt to capture the cognitive process underlying context effects on memory. Amongst those, a few offer potential explanations for how familiarity may influence the CR effect.

The "outshining hypothesis" (Smith, 1988, 1994; Smith & Vela, 2001) suggests that the target and the context are stored as separate entities which act as cues for one another. When the strength of the target item is weak relative to the strength of the context cue, a benefit of reinstating context will be observed. In other words, the degree to which context will act as a cue for a particular memory depends on the relative strength of the target in memory. Thus, this theory predicts that the benefit of reinstating context will vary; specifically, it will be greater for unfamiliar than for familiar target stimuli. As familiarity with the target increases, the outshining hypothesis suggests that context-dependent effects will be attenuated. That is, highly familiar relative to unfamiliar target stimuli may provide such strong memorability cues on their own that the memory trace for these items "outshines" any influence that the context might have on their memorability. Dalton (1993) provided empirical evidence for the outshining hypothesis, demonstrating a recognition benefit for more weakly encoded items (novel faces) that were tested in their original context (Room A) versus in a new context (Room B), but no such context benefit for faces familiarized by a single pre-study exposure, which occurred 1 week prior. This work was later replicated and extended to familiar and unfamiliar words (Isarida et al., 2012; Russo et al., 1999).

Another theory predicts a very different pattern of results. In contrast to the outshining hypothesis, global matching models (CHARM: Eich, 1982; TODAM2: Murdock, 1997) suggest that recognition decisions are determined by a sum of target and context signal strengths. When viewing a target-context pairing at test, the target and the context each produce contributions informing the recognition decision. The greater the match between stimuli presented during recognition (target + context) to the items stored in memory from encoding, the greater the contributions and, consequently, the greater the likelihood of a correct recognition response. Thus, similar to the outshining hypothesis, global matching models predict that context will benefit recognition of unfamiliar targets. However, when familiarity of the target is considered, very different predictions emerge. In contrast to the outshining hypothesis, in which the benefit of the context is predicted to be modulated by target familiarity, global matching models predict that context effects should be similar in magnitude *regardless* of target familiarity since target and context are assessed independently. That is, the model suggests that degree of match between the test probe (probe target + context) and a memory representation (amalgamated context + target unit) is an additive function of their featural overlap. The target and the context then do not act as cues for one another, but instead the memory for the *target* is matched to the probe *target*, while the memory for the *context* is independently matched to the probe *context*.

Support for global matching models has been found in a series of experiments conducted by Dougal and Rotello (1999). They examined the influence of familiarity of to-be-remembered words on context effects in a series of three experiments. Familiarity of target words was manipulated by increasing the frequency of presentation during encoding (Experiments 1 and 2) and through a levels-of-processing manipulation during encoding (Experiment 3), with different background colours serving as context. They found no influence of target familiarity on context effects. Note, however, that they did show a reduced CR effect for deeply relative to shallowly encoded words in Experiment 3, suggesting a diminishing effect of context as depth of processing increased. Prior to their study, and as already mentioned, Murnane and Phelps (1995) also found that manipulating familiarity, or cue strength, via number of exposures, length of presentation time, and levels-of-processing consistently produced the expected context reinstatement effects. Unfortunately, stimulus materials in both of these studies were confined to words, which, arguably are difficult to control for baseline familiarity (Tulving, 1972).

Other theories can be interpreted as making similar predictions. For example, a more recent model suggests that item-context pairings are encoded as single ensembles. The ICE model (Murnane & Phelps, 1999) suggests that information is stored as three separate types. *Item information* is defined as information that is central or primary to the performed task (i.e., target to be remembered item). *Context information* is any information that is peripheral or incidental to the primary cognitive task (i.e., the context scene). *Ensemble information* is information created when combining item and context information. In this theory, ensemble information produces greatest recognition performance when the item and context are the same at study and test. Since both the item and context information are required for the creation of ensemble representations in memory, context information is always important in determining recognition decisions. Therefore, similar to global matching models, the ICE model suggests that context should benefit item recognition memory regardless of item familiarity.

1.4 Overview of Experiments

In Chapter 2, I sought to investigate the impact of familiarity on the CR effect. Moreover, as previously discussed, there are a number of studies that have demonstrated the CR effect using various study stimulus types, including words, objects and faces. As such, I decided to investigate whether familiarity with each of these stimulus types would have a unique effect on the benefit of context reinstatement (Lee & Fernandes, submitted). In addition, no study to date has include more than a binary comparison of familiarity (i.e. famous vs. non-famous faces). In each of the experiments in Chapter 2, four pre-exposure conditions were included to map out how increasing familiarity can gradually influence the CR benefit. In Experiments 1, 2 and 3, familiarity of words, objects and faces, respectively, were manipulated by pre-exposing unfamiliar stimuli either 0, 1, 3 or 10 times. The purpose of Experiment 4 was to rule out any ceiling effects that may have explained the pattern observed in the previous three experiments.

The goal of Chapter 3 was to investigate a potential explanation for the pattern of results observed in Chapter 2. One such explanation is an attentional capture account. Participants completed a paradigm similar to that used in Experiments 1-4, but memory was assessed for target words and faces in separate blocks in a within-subjects design. Prior to each block, eye gaze was calibrated with an eye tracker as an indirect measure of overt attention.

In Chapter 4, I investigated how semantic characteristics of the context itself, along with individual differences across participants, could impact CR. There is some suggestion that the binding between an item and its context is weakened by strong emotionality, such that the ability of a context to support memory for its paired item would be reduced. In Chapter 4 I manipulated the anxiety-provoking nature of the context scene paired with target faces during encoding, and assessed the relative benefit of reinstating these contexts again during the recognition test. Moreover, I compared these effects across individuals endorsing high versus low levels of social anxiety in their daily lives.

My PhD experiments suggest a refinement to classic CR effect is needed: it applies reliably and primarily to memory for words or objects, and for faces that are relatively novel. When faces are familiar, or encoded in highly anxiety-provoking scenes, other factors, such as target signal strength are more important in guiding recognition memory decisions.

Chapter 2: Faces are Special: Familiar Faces, But Not Words or Objects, Reduce CR Benefit

One possible reason for the variable findings with respect to the influence of familiarity on context effects in the previously described research may lie in the choice of stimulus materials, notably faces versus words. As previously discussed, one of the main inconsistencies in reports of the context reinstatement effect is the variety of stimulus types that are studied and tested: they could be words (Isarida et al., 2012), objects (Craik & Schloerscheidt, 2011) or faces (Gruppuso et al., 2007, Watkins, Ho & Tulving, 1976). Additionally, work on eyewitness testimony has shown that participants are more likely to correctly identify target faces when provided with the same physical cues (Krafka & Penrod, 1985) or when returned to the same environmental context (Smith & Vela, 1992). A major disadvantage to using words (as targets) is that they are all generally familiar to participants, and it is therefore difficult to create a truly "unfamiliar" set of words (Tulving, 1972). The same may be argued for objects, considering at least a variant of many objects typically used in studies of this nature have been utilized by participants outside the lab in the past (i.e. a pencil). Using a set of face photographs, as in the current paradigm (Experiments 3 and 4), allowed us to create stimulus sets for which degree of familiarity could be controlled experimentally. Given that theories predict different effects of context depending on target familiarity, I sought clarification by using stimuli for which I could more easily determine levels of familiarity, to examine how this factor interacts with context effects.

In the current study, I investigated how familiarity with several kinds of stimuli (words, objects, and faces) influenced the degree to which context could benefit target memory. I manipulated familiarity using a pre-exposure phase during which target words (Experiment 1), objects (Experiment 2), or faces (Experiment 3) were presented either 0, 1, 3, or 10 times. In a fourth experiment, the 10 times pre-exposure condition for faces was reduced to a 5 times pre-exposure condition (Experiment 4). In each experiment, participants viewed target stimuli during an encoding phase, paired with indoor or outdoor pictures as context. At retrieval, targets and lures were presented or each trial of a recognition test, paired either with the same or with a new context.

I opted to include a range of pre-exposures so that I could track how context effects changed systematically, with increasingly familiar stimuli. The current design thus allowed for a more quantifiable assessment of the role of target familiarity on context effects, as opposed to the qualitative categorical separation used by studies in the past (Isarida et al., 2012; Russo et al., 1999; Reder et al., 2013). Furthermore, extending the findings of studies that did use a quantifiable manipulation of familiarity (Dalton, 1993; Murnane & Phelps, 1995; Schonbach-Medina & Vakil, 2017), the inclusion of a range of pre-exposure conditions (0, 1, 3, and 10) allowed us to map out and determine the impact of gradually increasing the degree of familiarity, from none to little to extremely familiar, on the CR effect.

Global matching models and the outshining hypothesis suggest two different outcomes for my experiments. According to global matching models, context should always contribute to recognition memory decisions. Therefore, we should always observe a contextual benefit on memory, regardless of target item familiarity. In contrast, according to the outshining hypothesis, memory for familiar stimuli may outshine memory for their context. Hence, for familiar stimuli (those viewed 10 times in the pre-exposure phase), CR effects (reinstating the same relative to a new context) should be reduced or completely eliminated relative to unfamiliar stimuli (those viewed 0 times in the pre-exposure condition).

2.1 Experiment 1

In Experiment 1, I investigated whether manipulating familiarity with target words, in word-scene pairs, would influence the benefit of context reinstatement. Familiarity was manipulated in a pre-exposure phase where words were viewed either 1, 3, or 10 times. Increasing the number of pre-exposures for a target word will correspondingly increase the strength of the memory signal for that word. As a result, according to the outshining hypothesis (Smith, 1988, 1994; Smith & Vela, 2001), memory for highly familiar words may outshine that for an associated context, thereby reducing the memory benefit of reinstating context. In terms of the design of this experiment, less familiar words (those not included in the pre-exposure phase) should therefore receive a greater benefit of context reinstatement than words pre-exposed 10 times.

A different outcome is suggested by global matching models (CHARM: Eich, 1982; TODAM2: Murdock, 1997) and the ICE model (Murnane & Phelps, 1999). These suggest that recognition decisions are determined by the sum of activated encodings for the target and context, or ensemble. As such, regardless of familiarity, or the number of pre-exposures, the presence of context will always contribute to recognition decisions and enhance word recognition. Though note that, as I expect target memory will be enhanced when number of pre-exposures is increased, there, theoretically, will be a point at which memory in both context conditions will be at ceiling. Once this limit is reached, there would be no CR benefit as memory has already reached its peak.

2.1.1 Method

Participants

Forty-four undergraduate students completed the study (36 females, *M* age = 19.80, *SD* = 2.12, *Range* = 18-28 years)[1]. Participants were enrolled in undergraduate Psychology classes and received course credit or token monetary remuneration for their participation. The mean number of years of education was 13.77 (*SD* = 1.67).

Materials

A set of 96 words was chosen from the Affective Norms for English Words (ANEW; Bradley & Lang, 1999) to serve as the targets. Words (e.g., table, kettle, basket) were selected to be concrete nouns that were rated as neutral in valence and arousal. The context stimuli consisted of 96 photographs of scenes (half indoor/half outdoor), presented in colour. Context scenes were collected through an Internet search of publicly accessible images. Scenes included images of outdoor (e.g., beach, park, basketball court) and indoor (e.g., living room, restaurant) scenery. Context scenes did not include images of faces.

Design and Procedure

The sets of 96 words and 96 context photos were randomly paired. The pairs were then randomly divided in half to create two study lists (List A and List B), each consisting of 48 words and 48 context scenes (half indoor/half outdoor) pairs. List A was presented to half of the

---

[1] A power analysis using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) was conducted assuming a small effect size (*d* = .27) as found in Isarida et al. (2012), and an alpha of .05. Results revealed a sample size of 20 would be required to achieve a power of .96. As this power calculation was conducted with a 2 x 2 design with only 2 familiarity conditions, I doubled the sample size in this study to correspond with the 4 familiarity levels.

participants during the study phase. Words from the second unused list (counterbalanced across participants) were used as lures during the test phase.

Prior to study, a pre-exposure phase was administered to familiarize participants with the words to be used in the study. Thirty-six of the 48 words from the predetermined study list (A or B) were divided into 3 groups. Of the 36 study words, 12 were presented to participants once (1 pre-exposure), 12 were shown 3 times (3 pre-exposures) and 12 were shown 10 times (10 pre-exposures). The remaining 12 words were not shown to participants. There were therefore 4 pre-exposure conditions that varied, within-participant, in terms of familiarity.

On each trial during the pre-exposure phase, participants viewed each word in black, size 32 font, presented alone on a white background, in the centre of the screen for 3500 msec, followed by a fixation cross for 500 msec. The set of words was presented with pre-exposure condition randomized. Participants were asked to press the “v” key on a standard QWERTY keyboard if the word contained the letter ‘a’, or the “b” key if the word did not contain ‘a.’ The purpose of this task was to encourage processing of the word on each trial. Past research has also indicated that the CR benefit is reduced when completing a deep encoding task (Dougal & Rotello, 1999). Therefore, to optimize context effects in the eventual test, I asked participants to complete a shallow encoding task.

Following the pre-exposure phase, in an incidental encoding phase, participants were presented with 48 word-context pairs from the predetermined list (A or B, counterbalanced across participants). The pairs were each presented centrally on the screen for 2250 msec on a white background. Each word-context pair was presented sequentially in a randomized order. A screen appeared after each pair displaying a Likert-type scale and remained on the screen until the participant made a response. Using the top row of numbers on a standard keyboard, participants rated the likelihood that the word would be associated with the accompanying context scene (1 = very unlikely, 6 = very likely). This association was made as there is some argument that context will not benefit memory without inclusion of a binding task (Humphreys & Chalmers, 2016). A sample of a study word paired beside a context scene is shown in Figure 1A.

Throughout the pre-exposure and encoding phases, it was not made evident to participants that they would be later tested on memory for the words. Instead, participants were specifically told to make decisions to associate the pairs together. In addition, the study was labeled, during recruitment, as an assessment of feelings of belonging, to further mask the upcoming recognition test.

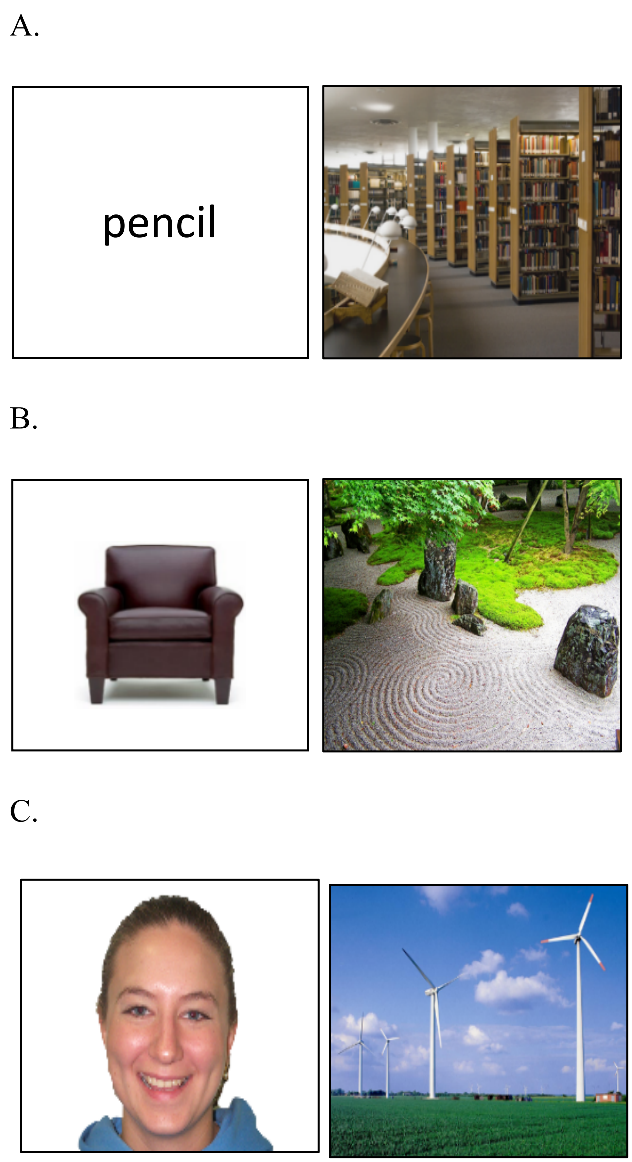

Figure 1. Sample of stimuli (A. word, B. object, C. face) paired alongside a context scene.

A recognition test was administered following the encoding phase. During each trial of the recognition test, a word-context pair was displayed for 5000 msec, and participants were instructed to make a recognition decision for the word. If they judged that the word had been seen previously during the encoding phase, they clicked the “1” key on the number pad, marked by a sticker with an “R”, signifying it was deemed to be remembered. To indicate the word was new (i.e., that they did not remember seeing it), participants were instructed to click the number “3” on the number pad, marked by an “N” signifying it was judged to be a new word.

At test, half of the 48 word-context pairs from the encoding phase remained intact and the other half were re-paired. The re-paired word-context pairs contained either a word or context scene shown previously during encoding, but alongside a novel word or novel context scene. Therefore, there were four test trial types: 1) old word + same context, 2) old word + new context, 3) new word + old context, 4) new word + new context. Target words were tested paired alongside novel contexts, as opposed to re-matched pairs, because it was important to test memory for the word itself, rather than for the pairing. In doing so, I could more accurately deduce how the presence of context influences the memory of a familiar, or unfamiliar word. In total, there were 48 target and 48 foil words presented during the recognition test; the same was true for the context pictures. Among the word-context pairs that included target words, half (24) remained intact from encoding, while the other 24 words were paired with novel context scenes. For the 48 foil words, half (24) were paired with contexts that had been previously seen during the encoding phase; the remaining 24 were paired with novel context scenes.

Study lists were counterbalanced across participants. That is, List A was used as the target study list and List B was used as the foil list for half of the participants, and this was reversed for the other half of the participants. For both study and test, the target was always presented on the left and the context scene on the right (see Figure 1).

2.1.2 Results

Accuracy Rate

Accuracy was calculated as hit rate minus false alarm rate, for each Pre-Exposure and Test Context condition. Across a number of studies, there is evidence that pairing lures with previously viewed contexts increases false alarm rate. Therefore, it was important to use a memory measure that accounts for both hit and false alarm rates. Hit rate was calculated as the number of correctly identified words that were previously presented at encoding divided by 6, in each pre-exposure condition. False alarm rate was calculated as the number of foil words that were endorsed as old divided by 24[2]. Mean hit rates, false alarm rates, and accuracy are summarized in Table 1.

To determine the effect of pre-exposure familiarization on word recognition, I conducted a 4 (Number of Pre-Exposures: 0, 1, 3, 10) × 2 (Test Context: Same or New) repeated-measures ANOVA. Results indicated that there was a main effect of Pre-Exposures $F(3, 129) = 4.98$, $MSE = 0.16$, $\eta_p^2 = .10$, $p < .01$, such that words pre-exposed 0 times were recognized significantly worse than words pre-exposed once, ($F(1, 43) = 7.50$, $MSE = 0.03$, $\eta_p^2 = .15$, $p < .01$), three times ($F(1, 43) = 12.32$, $MSE = 0.03$, $\eta_p^2 = .10$, $p = .001$), and 10 times ($F(1, 43) = 8.30$, $MSE = 0.04$, $\eta_p^2 = .16$, $p < .01$). The main effect of Test Context was also significant, $F(1, 43) = 34.71$, $MSE = .13$, $\eta_p^2 = .45$, $p < .001$. Words presented with the same context as at encoding were

[2] Note that false alarm rates did not differ for lures paired with same or new context scenes ($p > .90$). Mean false alarm rates are shown in Table 1.

recognized more often than words re-paired with a new context (M_{same} = .89, M_{new} = .66). The Pre-Exposures × Test Context interaction was not significant, $p > .05$[3].

Comparing magnitude of context reinstatement

I compared the magnitude of the CR effect for words in each pre-exposure condition (1, 3, and 10 times) relative to the 0 pre-exposure condition. Magnitudes (shown in Figure 2) were calculated as the difference in accuracy between words paired with the same versus new contexts. I conducted a one-way ANOVA to determine whether the CR magnitude differed depending on number of pre-exposures. Results showed no main effect of Pre-Exposure condition on CR magnitude, $F(3, 129) = 1.32$, $MSE = .03$, $\eta_p^2 = .03$, $p = .27$.

---

[3] Across all four experiments, patterns of results for hit rates were the same as the patterns found for hit minus false alarm accuracy.

Table 1

Experiments 1-4: hit rate, false alarm rate and accuracy rate for each material type, pre-exposure condition, and recognition test context (standard deviation in parentheses).

	Context at Test	Hit Rate (SD)	ACC (SD)	Hit Rate (SD)	ACC (SD)	Hit Rate (SD)	ACC (SD)	Hit Rate (SD)	ACC (SD)	FA Rate (SD)
Words		0 Exposures		1 Exposure		3 Exposures		10 Exposures		
	Same	.91 (.13)	.86 (.15)	.96 (.08)	.91 (.13)	.94 (.10)	.89 (.12)	.96 (.08)	.91 (.09)	.06 (.07)
	New	.66 (.26)	.60 (.25)	.72 (.30)	.66 (.29)	.76 (.29)	.70 (.28)	.73 (.32)	.67 (.31)	.05 (.06)
Objects		0 Exposures		1 Exposure		3 Exposures		10 Exposures		
	Same	.89(.16)	.86 (.17)	.93 (.15)	.90 (.17)	.98 (.06)	.95 (.10)	.96 (.09)	.93 (.10)	.03 (.01)
	New	.37(.29)	.26 (.27)	.55 (.37)	.45 (.33)	.51 (.40)	.41 (.38)	.57 (.42)	.47 (.39)	.10 (.02)
Faces		0 Exposures		1 Exposure		3 Exposures		10 Exposures		
	Same	.59 (.23)	.51 (.25)	.80 (.19)	.72 (.21)	.94 (.11)	.87 (.16)	.96 (.10)	.88 (.17)	.08 (.10)
	New	.47 (.22)	.42 (.22)	.70 (.23)	.65 (.24)	.82 (.20)	.77 (.21)	.94 (.13)	.89 (.15)	.05 (.07)
Faces (Experiment 4)		0 Exposures		1 Exposure		3 Exposures		5 Exposures		
	Same	.69 (.25)	.59 (.26)	.83 (.17)	.73 (.20)	.97 (.08)	.87 (.14)	.96 (.08)	.86 (.13)	.08 (.10)
	New	.47 (.24)	.40 (.23)	.67 (.26)	.60 (.28)	.83 (.23)	.76 (.23)	.88 (.23)	.82 (.25)	.07 (.09)

Note: ACC = accuracy rate, FA = false alarm, SD = standard deviation

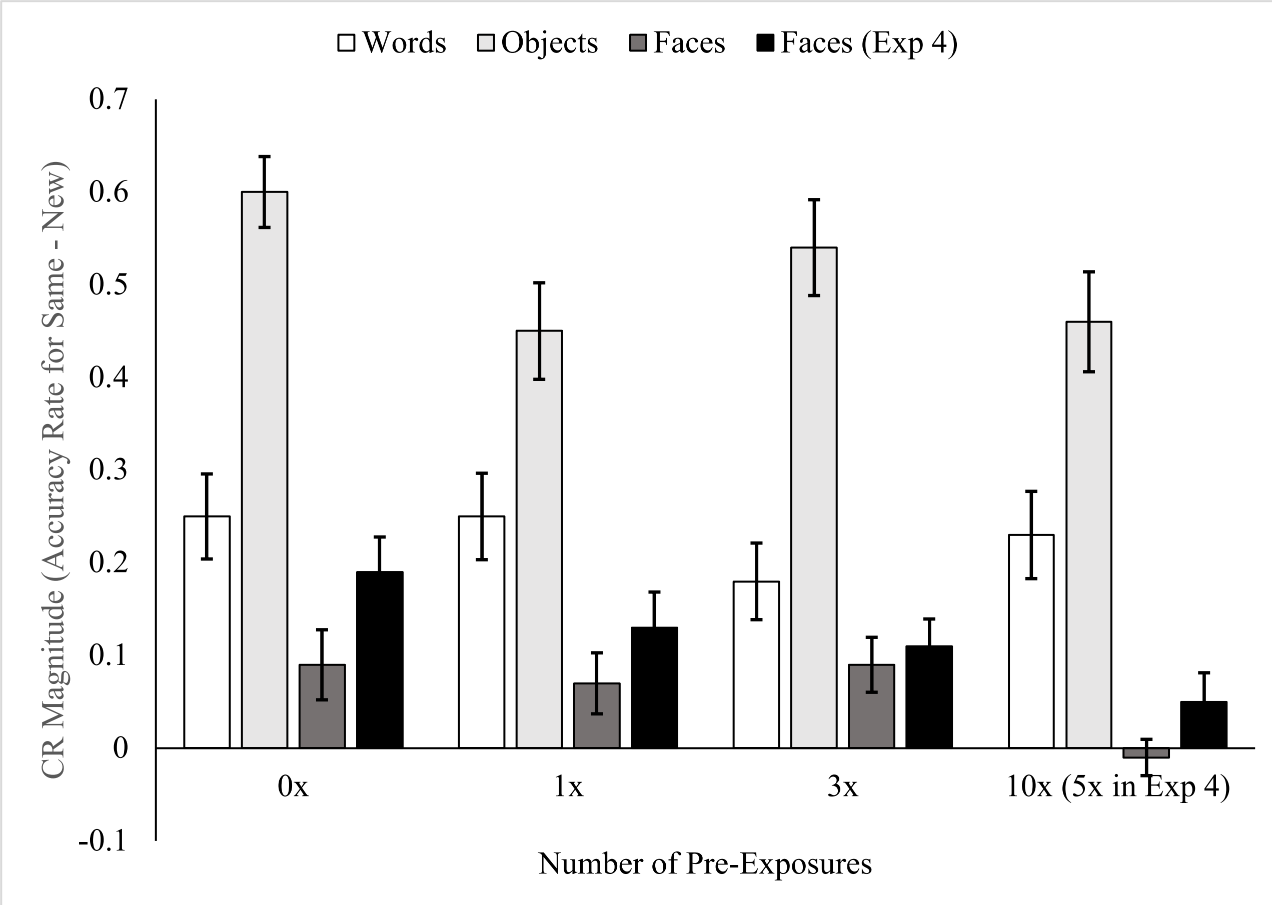

Figure 2. Magnitude of context reinstatement effect (recognition accuracy rate for same minus new context) for various stimulus types, as a function of number of pre-exposures (error bars show standard error of the mean).

Relation between subjective association rating at encoding and CR magnitude

There is some evidence to suggest that the ease with which a target and the associated context are bound influences the magnitude of the CR effect. That is, Reder et al, (2013) have suggested that highly familiar targets are more easily bound to contexts because they have pre-existing representations, and that this enhances CR effects. To examine this possibility in the data, I correlated the subjective ratings of degree of association between target and context, made by participants during encoding, with the CR magnitude, within each pre-exposure condition. Mean subjective association ratings for all four experiments are shown in Table 2. Correlations between these and CR magnitude were not significant in any of the pre-exposure conditions: 0 times ($r(44) = -.05$, $p = .73$), 1 time ($r(44) = -.22$, $p = .16$), 3 times ($r(44) = -.21$, $p = .73$), or 10 times ($r(44) = -.09$, $p = .59$).

Table 2

Experiments 1-4: mean subjective association ratings for each stimulus type and pre-exposure condition (standard deviation in parentheses).

Stimulus Type	0 Pre-Exposures	1 Pre-Exposure	3 Pre-Exposures	10 Pre-Exposures (5 for Experiment 4)
Words	2.03 (.63)	1.92 (.57)	2.29 (.69)	2.16 (.61)
Objects	2.38 (.70)	2.04 (.44)	2.85 (.68)	2.54 (.43)
Faces	3.74 (.68)	4.02 (.66)	3.99 (.59)	4.21 (.55)
Faces (Experiment 4)	3.86 (.74)	3.97 (.76)	3.92 (.69)	4.02 (.72)

2.1.3 Discussion

In this experiment, I was able to demonstrate the CR effect for words (Hockley, 2008; Murnane & Phelps, 1993, 1994, 1995; Smith & Vela, 2001). There was, however, no effect of pre-exposure condition on CR magnitude. This pattern of results is in line with the prediction made by global matching (CHARM: Eich, 1982; TODAM2: Murdock, 1997) and ICE (Murnane & Phelps, 1999) models. That is, recognition performance improved when study contexts were re-presented during retrieval, regardless of one's familiarity with the target. This pattern of results is in contrast to the predictions made by the outshining hypothesis (Smith & Vela, 2001), which predicts that CR magnitude should decline as target familiarity increases. There were also no correlations between subjective ratings of association for targets to contexts and magnitude of CR effect, in any of the pre-exposure conditions. Thus, increasing familiarity, via number of pre-exposures with a word, did not increase the magnitude of the CR benefit.

In Experiment 1, I found no influence of manipulated familiarity on the magnitude of the CR effect. However, as discussed briefly in the introduction, it may be that the target stimulus type matters, as the CR effect varied considerably depending on whether the target was a word, non-word, or face. In my next experiments, I systematically varied the target material to determine whether the absence of an effect of pre-exposure (and familiarity) would influence the magnitude of CR in the same way for all materials. That is, I looked for converging evidence to support global matching and ICE models, in explaining the influence of familiarity on the CR benefit to memory.

2.2 Experiment 2

In Experiment 1, I found that familiarity with target words did not influence the magnitude of the CR effect. Past evidence suggesting that target familiarity influences CR effects has been mixed (Isarida et al., 2012; Murnane & Phelps, 1995; Reder et al., 2013; Russo et al., 1999; Schonbach-Medina & Vakil, 2017). A potential problem with studies using sets of words, including in my experiment, is that the words are never truly unfamiliar to participants (see also Tulving, 1972, for a similar claim). Thus, my ability to make strong conclusions about the role of familiarity based on Experiment 1 is limited.

In Experiment 2, I replaced the target stimulus set of words with a set of objects, to investigate whether previously unseen, and thus unfamiliar, objects may yield a different pattern of results. There is evidence of a CR benefit for objects (Craik & Schloerscheidt, 2011). Given that I used picture stimuli as the context scenes, it may be easier to bind object pictures to the scenes in comparison to, arguably, visually unrelated word stimuli. Such binding ease may allow for an effect of familiarity, as suggested by others (Reder et al., 2013). As a result, I predicted I would observe an enhancement in memory corresponding with an increase in pre-exposures. Global matching and ICE models do not differentiate between stimulus types; the prediction from these is for a replication of the results of Experiment 1. Conversely, the outshining hypothesis predicts a reduction in the CR effect for familiar objects, pre-exposed 10 times, relative to unfamiliar objects.

2.2.1 Method

Participants

Fifty-two undergraduate students completed the study (36 females, *M* age = 19.83, *SD* = 1.95, *Range* = 17-25 years). Participants were enrolled in undergraduate Psychology classes and received course credit or token monetary remuneration for their participation. The mean number of years of education was 13.83 (SD = 1.78).

Materials

The objects were chosen from a database used in previous work testing recognition memory for a set of objects (Meade & Fernandes, 2017). Because participants needed to make binary decisions during the pre-exposure phase of this study, it was important to select objects that would easily fit into two categories (i.e., living or non-living). All objects were concrete items, consisting of either living animals or household items and were displayed in colour, with dimensions of 512 x 384 pixels. A stimulus sample paired beside a context scene is shown in Figure 1B. The same set of context scene stimuli from Experiment 1 was used in Experiment 2.

Design and Procedure

The same procedure was applied here as in Experiment 1. The critical difference was that objects were used as the to-be-remembered target stimuli instead of words. During the pre-exposure phase, participants were asked to identify whether the object was living or non-living by making a keypress. The purpose of this task was to encourage processing of the object on each trial.

The remaining segments of the procedure remained the same as in Experiment 1.

2.2.2 Results

Accuracy Rate

Accuracy was calculated as the difference between hits and false alarms for each Pre-Exposure and Test Context condition (see Table 1). To determine the effect of Pre-Exposure on object recognition accuracy, I conducted a 4 (Number of Pre-Exposures: 0, 1, 3, 10) × 2 (Test Context: Same or New) repeated-measures ANOVA. Greenhouse-Geisser corrections were applied to correct for a violation of sphericity. Results revealed a main effect of Pre-Exposures, $F(2.58, 131.72) = 24.23$, $MSE = 0.02$, $\eta_p{}^2 = .32$, $p < .001$. Accuracy was significantly lower for objects in the 0 pre-exposure condition compared to all other conditions (1 pre-exposure: $F(1, 51) = 42.24$, $MSE = .03$, $\eta_p^2 = .45$, $p < .001$; 3 pre-exposures: $F(1, 51) = 29.51$, $MSE = .05$, $\eta^2_p = .37$, $p < .001$; 10 pre-exposures: $F(1, 51) = 50.04$, $MSE = .04$, $\eta_p^2 = .50$, $p < .001$). Accuracy for objects in the 1, 3, and 10 pre-exposure conditions did not differ from each other. There was also a significant main effect of Test Context, $F(1, 51) = 145.29$, $MSE = .19$, $\eta_p^2 = .74$, $p < .001$. Objects presented with the same context as at encoding were better recognized than objects re-paired with a new context ($M_{\text{same}} = .91$, $M_{\text{new}} = .40$).

The Pre-Exposures × Test Context interaction was significant, $F(3, 153) = 5.94$, $MSE = .02$, $\eta_p^2 = .10$, $p = .001$. Importantly, however, a priori planned analyses of pre-exposure level contrasts revealed accuracy was significantly enhanced in each pre-exposure condition. That is, for objects in the 0 ($t(51) = 15.68$, $p < .001$), 1 ($t(51) = 8.67$, $p < .001$), 3 ($t(51) = 10.48$, $p < .001$), and 10 ($t(51) = 8.50$, $p < .001$) Pre-Exposure conditions, accuracy was significantly higher when context was the Same, as opposed to New[4].

Comparing magnitude of context reinstatement

I compared the magnitude of the CR effect for objects in each pre-exposure condition (1, 3, and 10) relative to the 0 pre-exposure condition. CR magnitudes are shown in Figure 2. I conducted a one-way ANOVA to determine whether the CR magnitude differed depending on number of pre-exposures. There was a main effect of Pre-Exposure condition on CR magnitude, $F(3, 153) = 5.94$, $MSE = .04$, $\eta_p^2 = .10$, $p < .001$. Simple effects contrasts revealed that the CR magnitudes for objects in the 10 ($M = .45$, $SD = .39$; $F(1, 51) = 8.92$, $MSE = .12$, $\eta_p^2 = .15$, $p <$

[4] Paired samples t-test revealed a significant difference in false alarms ($t(51) = 4.10$, $p < .001$), whereby there were significantly more false alarms to lures paired with new ($M = .10$, $SD = .12$), compared to old ($M = .03$, $SD = .05$) context scenes. Despite this difference in false alarms, though, the pattern of hits was the same as that for accuracy.

.01) and the 1 (M = 45, SD = .38; $F(1, 51) = 10.63$, MSE = .11, $\eta^2_p = .17$, $p < .01$) pre-exposure conditions were significantly lower than the magnitude for objects in the 0 (M = .60, SD = .28) condition. There was no difference in CR magnitude between objects in the 3 (M = .54, SD = .37) and 0 conditions, $F(1, 51) = 1.68$, MSE = .10, $\eta_p{}^2 = .03$, $p = .20$.

Relation between subjective association rating at encoding and CR magnitude

As in Experiment 1, we conducted Pearson correlations to examine the relation between CR magnitude and subjective association ratings at encoding. Correlations between mean subjective association ratings for objects and CR magnitude were not significant in the 0 ($r(52) = -.09$, $p = .51$), 1 ($r(52) = .02$, $p = .89$), 3 ($r(52) = .02$, $p = .88$), or 10 ($r(52) = .07$, $p = .64$) pre-exposure conditions. There is thus no relation between subjective association and CR magnitude.

2.2.3 Discussion

In this experiment, I chose to use objects instead of words as the target stimuli. Although I was able to find a significant interaction between test context and the number of pre-exposures to the target objects, contrary to my predictions, I did not find that increasing the degree of familiarity (by increasing number of pre-exposures) further influenced the CR effect. That is, objects pre-exposed once produced recognition performance rates similar to those of objects pre-exposed 3 and 10 times. CR magnitude, however, was higher for objects in the 0 pre-exposure condition than in the 1 and 10 conditions. This reduction in CR benefit when objects were pre-exposed 1 and 10 times suggests that perhaps pre-exposure does play some role in influencing the CR effect.

The results of this experiment provide mixed support for both the outshining hypothesis and global matching/ICE models. In line with the outshining hypothesis, CR magnitude was greatest for the most unfamiliar objects (0 pre-exposures). However, further increasing the degree of familiarity from 1 to 10 times did not systematically reduce the CR effect, contrary to the prediction of the outshining hypothesis. One possible explanation is that a single exposure to a target test stimulus is sufficient to make an item (at least an object) familiar (Jacoby, Kelley, Brown & Jasechko, 1989; Jacoby, Woloshyn, & Kelley, 1989). As such, increasing the number of pre-exposures any further would not provide any additional influence. The reduction in CR effect, may have already reached its peak after a single pre-exposure. It is possible, however, that memory performance was at ceiling, making it difficult to discern any benefit of context. Though memory performance was quite high, this great performance was only observed when context was reinstated. I anticipated that, with increasing number of pre-exposures, memory performance for the objects would improve in both test context conditions. This was not the case. Instead, memory performance when paired with a novel context remained consistent across the 1, 3 and 10 pre-exposure conditions. Across the four pre-exposure conditions, and in line with global matching models, I found evidence that context reinstatement consistently benefitted memory performance.

Given these findings, it does not appear that increasing familiarity of targets reliably reduces CR effects. As well, as in Experiment 1, there was again no correlation between CR magnitude and subjective rating given at encoding of match between target and context. In addition, subjective rating scores ranged from 1 to 4.5, spanning most of the 6-point Likert scale. CR magnitude ranged from -.42 to 1, so there is most likely no issue with restriction of range. This suggests that the CR effect is reliable and robust, regardless of the perceived semantic relation between target and context. This conclusion is counter to the claim by Reder and colleagues (2013), who suggest that CR effects are dependent on qualities of the target and its 'bindability' with context.

In their study, during encoding, famous and non-famous faces were superimposed on various familiar context scenes (e.g. the Eiffel Tower). In a surprise recognition test, old and new faces were paired with either a reinstated context scene, or one originally paired with a different face (recombined). Memory performance was higher for famous faces paired with a reinstated, compared to a recombined, context. This pattern was not observed for non-famous faces. The authors proposed that it is easier to associate context to faces that have a pre-existing long-term memory representation (famous ones) than to faces that do not (unfamiliar ones). However, recall that there also exists evidence to the contrary. Dalton (1993) showed that there is a reduction in the CR effect when targets are familiar faces. Given the mixed findings, in my next experiment, I investigated whether and how familiarity with faces, using my pre-exposures paradigm, influenced the context reinstatement effect.

2.3 Experiment 3

Evidence of the influence of familiarity on CR effects for target faces has been mixed. Along the same lines as Murnane and Phelps (1995), Reder and colleagues (2013) found evidence that the CR effect is particularly beneficial for faces with pre-existing memory representations (i.e. famous faces). Experience, however, tells us that there are certain faces we are able to easily recognize in any context, regardless of context (for example, one's supervisor or one's hockey coach). A drawback of the study by Reder and colleagues (2013) is that the degree to which each participant was familiar with each famous face likely varied. Similar to words, this variability in baseline familiarity makes it difficult to create a truly familiar set of famous faces. In my own pilot study, the rated knowledge of a set of famous faces varied considerably (from 41% to 100%). Indeed, previous research has found that even one presentation of a target stimulus, such as a name, can make a stimulus 'famous' (Jacoby et al., 1989a, 1989b). Furthermore, evidence has shown that the CR effect diminished following even a single pre-exposure (Dalton, 1993). However, no study to date has determined whether changing the *degree* of familiarity with a face influences context effects. This is important as, to this point, studies have focused on binary manipulations of familiarity (i.e. famous and non-famous). Quantifying and manipulating the degree of familiarity makes it possible to determine the point at which a face switches from unfamiliar to 'famous'. In Experiment 3, I was able to control for the degree of familiarity, or 'fame' status using my pre-exposure paradigm.

In Experiment 3, I used a set of unfamiliar faces in my pre-exposure paradigm to determine whether manipulating the degree of familiarity with a target face would reduce the CR effect (Dalton, 1993), or enhance it (Reder et al., 2013). Such a pattern of results would be in line with the outshining hypothesis. In contrast, and in accordance with my previous two experiments and global matching/ICE models, familiarity may not reduce the CR magnitude. Instead, we may observe a benefit of context regardless of familiarity.

2.3.1 Method

Participants

Fifty-four undergraduate students completed the study. Data for 5 participants were excluded as they failed to correctly follow instructions (i.e., pressing the incorrect key to indicate an 'old' response) and 2 were removed due to excessively low accuracy rates (three standard deviations below the mean). Of the remaining 47 participants, 38 were females. Participants had a mean age of 19.63 (*SD* = 3.15, *Range* = 18-35 years), and 13.07 (*SD* = 1.46) years of education. All participants were enrolled in undergraduate Psychology classes and received course credit or token monetary remuneration for their participation.

Materials

Faces were chosen from the Center for Vital Longevity Face Database (Minear & Park, 2004). Faces included those of young to middle-aged adults from an assortment of ethnicities to match the heterogeneity of the sampled student population. Photographs included head and shoulders only; Faces contained no distinguishing features (e.g. facial accessories, glasses, sunglasses, hats), and were presented in front view, in colour, and on a white background (see Figure 1C). The same set of context scene stimuli from the previous experiments was used in Experiment 3.

Design and Procedure

A similar procedure to that of Experiments 1 and 2 was used in Experiment 3. The key difference was that the words and objects from Experiment 1 and 2, respectively, were replaced with images of faces. During the pre-exposure phase, participants were asked to report whether the face was of a male or female by pressing the ‘v’ key to indicate male and the ‘b’ key to indicate female. Again, the purpose of this task was to ensure processing of the face on each trial.

The remaining segments of the procedure was unchanged.

2.3.2 Results

Accuracy Rates

Accuracy was calculated as the difference between hits and false alarms, for each Pre-Exposure and Test Context condition (see Table 1). To determine the effect of Pre-Exposures and Test Context on face recognition accuracy, I conducted a 4 (Pre-Exposures: 0, 1, 3, 10) × 2 (Test Context: Same or New) repeated-measures ANOVA. After applying Greenhouse-Geisser corrections, the analysis revealed a main effect of Pre-Exposure, $F(2.41, 110.76) = 107.02$, $MSE = 0.04$, $\eta_p^2 = .70$, $p < .001$. As anticipated, as number of pre-exposures increased, so too did recognition accuracy for target faces (M_0 = .47, M_1 = .69, M_3 = .82, M_{10} = .89). Faces pre-exposed once were recognized significantly better than faces in the 0 pre-exposure condition ($F(1, 46) = 52.27$, $MSE = .09$, $\eta_p^2 = .53$, $p < .001$); faces in the 3 condition were recognized significantly more often than those in the 1 condition ($F(1, 46) = 27.70$, $MSE = .06$, $\eta_p^2 = .38$, $p < .001$); and accuracy was significantly higher for faces pre-exposed 10 times than 3 times ($F(1, 46) = 20.40$, $MSE = .02$, $\eta_p^2 = .31$, $p < .001$).

The effect of Test Context was also significant, $F(1, 46) = 10.68$, $MSE = .03$, $\eta_p^2 = .19$, $p < .01$, demonstrating the CR effect (M_{same} = .75, M_{new} = .69). There was also a significant Pre-Exposure × Test Context interaction, even after applying the Greenhouse-Geisser correction, $F(2.65, 121.69) = 3.14$, $MSE = .02$, $\eta_p^2 = .06$, $p < .05$. Specifically, as revealed by a priori planned contrasts, the difference in accuracy between Same and New trial types was lowest, and non-significant, for faces pre-exposed 10 times (M_{same} = .88; M_{new} = .89; $t(46) = .68$, $p = .50$) and was highest for faces pre-exposed 0 times (M_{same} = .51; M_{new} = .42; $t(46) = 2.38$, $p < .05$). As expected, though significant, the difference in accuracy for faces pre-exposed 3 times (M_{same} = .87; M_{new} = .77; $t(46) = 3.15$, $p < .01$) and 1× (M_{same} = .72; M_{new} = .65; $t(46) = 2.19$, $p < .05$) was intermediate between the faces pre-exposed the most and fewest times[5].

Comparing magnitude of context reinstatement

I compared the magnitude of the CR effect for faces in each pre-exposure condition (1, 3, and 10) relative to the 0 pre-exposure condition. CR magnitudes are shown in Figure 2. I conducted a one-way ANOVA to determine whether the CR magnitude differed depending on

---

[5] Paired samples t-test revealed a significant difference in false alarm rates ($t(46) = 2.32$, $p < .05$), whereby there were significantly more false alarms to lures paired with old ($M = .08$, $SD = .10$), compared to new ($M = .05$, $SD = .07$) context scenes. Despite this difference in false alarm rates, though, the pattern for hits was the same as that for accuracy.

number of pre-exposures. Greenhouse-Geisser corrections were applied to correct for a violation of sphericity. Results revealed a main effect of Pre-Exposure condition on CR magnitude, $F(2.65, 121.69) = 3.14$, $MSE = .04$, $\eta_p^2 = .06$, $p < .05$. Planned simple effects contrasts revealed that the CR magnitude for faces in the 10 ($M = -.01$, $SD = .13$) condition was significantly lower than the magnitude for faces in the 0 ($M = .09$, $SD = .26$) condition, $F(1, 46) = 6.91$, $MSE = .07$, $\eta_p^2 = .13$, $p < .05$. All other contrasts were non-significant.

Relation between subjective association rating at encoding and CR magnitude

As in the previous experiment, I conducted Pearson correlations between CR magnitude and subjective ratings of degree of match between targets and context, done during encoding. There was a significant correlation for faces in the 3 pre-exposure condition ($r(47) = .29$, $p < .05$). However, correlations between mean subjective association ratings for faces and CR magnitude were not significant in the 0 ($r(47) = .16$, $p = .27$), 1× ($r(47) = -.18$, $p = .24$), and 10 ($r(47) = -.09$, $p = .57$) pre-exposure conditions.

2.3.3 Discussion

I found that increasing the number of pre-exposures with a target face reduced the benefit of CR. That is, the difference in memory performance, when retrieval context was the same as at encoding compared to new, was greatest for the least familiar faces (i.e., those not included in the pre-exposure phase). On the other hand, the difference in performance (across Same and New context reinstatement trials) was smallest for faces that were most familiar, those in the 10 pre-exposure phase. The CR magnitudes for faces in the 1 and 3 pre-exposure conditions were larger than for faces in the 0 condition and smaller than for faces in the 10 condition. This pattern suggests that the magnitude of the CR benefit is related to the degree of familiarity, or number of pre-exposures, to a target face.

One issue with the current study is that memory performance, specifically in the 10 pre-exposure condition, was near ceiling. As such, any potential benefit of context reinstatement, for highly familiar faces in that condition, may have been limited. To address the potential issue of a ceiling effect, in Experiment 4, I replaced the 10 pre-exposure condition with a 5 pre-exposure condition.

2.4 Experiment 4

Experiment 4 was identical to Experiment 3, except that the 10 pre-exposure condition was changed to 5. Here I attempted to replicate the findings of Experiment 3, such that the benefit of CR gradually decreased as familiarity with the target face increased. I expected that the results of this experiment would replicate those of Experiment 3, with the key difference that memory for the most familiar faces (5 pre-exposures) would no longer be at ceiling. As indicated by the results of Experiment 3, participant recognition of faces improved with increasing number of pre-exposures. It is therefore reasonable to suggest that we may observe a relatively decreased recognition performance in a 5, compared to a 10, pre-exposure condition.

2.4.1 Method

Participants

Fifty-one undergraduate students completed the study. Data from one participant were excluded due to excessively low accuracy (three standard deviations below the mean). Among the remaining fifty participants, 42 were female. Participants had a mean age of 20.12 (SD = 2.85, *Range* = 17-32 years) and 13.67 (SD = 1.41) years of education. Participants were enrolled in undergraduate Psychology classes and received course credit or token monetary remuneration for their participation.

Materials

The faces used in Experiment 4 were the same as those used in Experiment 3. The same set of context scene stimuli from the previous experiments was used in Experiment 4.

Design and Procedure

The procedure for Experiment 4 was similar to that of Experiment 3. In an attempt to reduce any potential ceiling effects, I replaced the 10 pre-exposure condition with a 5 pre-exposure condition. The other conditions remained the same as previous experiments.

2.4.2 Results

Accuracy Rate

Accuracy was calculated as the difference between hits and false alarms, for each Pre-Exposure and Test Context condition (see Table 1). To determine the effect of Pre-Exposure and Test Context on face recognition accuracy rate, I conducted a 4 (Number of Pre-Exposures: 0, 1, 3, 5) × 2 (Test Context: Same or New) repeated-measures ANOVA. Greenhouse-Geisser corrections were applied to correct for a violation of sphericity. Results revealed a main effect of Pre-Exposure, $F(2.19, 107.11) = 88.05$, $MSE = .04$, $\eta_p{}^2 = .64$, $p < .001$. As the number of pre-exposures increased, accuracy for target faces also increased. Simple effects contrast revealed that accuracy for faces in the 1 pre-exposure ($M = .67$) condition was higher than accuracy for faces that were not pre-exposed ($M = .50$; $F(1, 49) = 45.44$, $MSE = .07$, $\eta_p{}^2 = .48$, $p < .001$). There was also a significant increase when comparing accuracy for faces pre-exposed 3 times ($M = .82$) and 1 time, $F(1, 49) = 52.25$, $MSE = .04$, $\eta_p{}^2 = .52$, $p < .001$. Accuracy for faces was only marginally higher in the 5 ($M = .84$) pre-exposure condition than in the 3 pre-exposure condition, $F(1, 49) = 3.04$, $MSE = .02$, $\eta_p{}^2 = .06$, $p = .09$, and was significantly higher than in the 1 condition ($F(1, 49) = 58.99$, $MSE = .05$, $\eta_p^2 = .55$, $p < .001$) and the 0 ($F(1, 49) = 138.12$, $MSE = .09$, $\eta_p^2 = .74$, $p < .001$) conditions.

There was also a significant main effect of Test Context, $F(1, 49) = 24.99$, $MSE = .06$, $\eta_p^2 = .34$, $p < .001$. Accuracy for faces paired with the same compared to new contexts was higher ($M_{same} = .76$, $M_{new} = .65$). The Pre-Exposure × Test Context interaction was also significant, $F(2.33, 113.94) = 4.09$, $MSE = .03$, $\eta_p^2 = .08$, $p < .05$. Critically, a priori planned contrasts revealed the difference in accuracy between Same and New trial types was significant for faces in the 0 ($M_{same} = .59$, $M_{new} = .40$; $t(49) = 5.01$, $p < .001$), 1 ($M_{same} = .73$, $M_{new} = .60$; $t(49) = 3.28$, $p < .01$), and 3 ($M_{same} = .87$, $M_{new} = .76$; $t(49) = 3.70$, $p < .01$) conditions, but not when faces were pre-exposed 5 times, ($M_{same} = .86$, $M_{new} = .82$; $t(49) = 1.46$, $p = .15$). This result provides evidence that, the CR benefit was completely eliminated for the most familiar faces. Additionally, the difference in accuracy between Same and New trial types was lowest for faces pre-exposed 5 times and was highest for faces pre-exposed 0 times[6].

Comparing magnitude of context reinstatement

I compared the magnitude of the CR effect for faces in each pre-exposure condition (1, 3, and 5) relative to the 0 pre-exposure condition. CR magnitudes are shown in Figure 2. I conducted a one-way ANOVA to determine whether the CR magnitude differed depending on number of pre-exposures. Greenhouse-Geisser corrections were applied to correct for a violation of sphericity. As in Experiment 3, results of the ANOVA replicated the main effect of Pre-Exposure condition on CR magnitude, $F(2.33, 113.94) = 4.09$, $MSE = .06$, $\eta_p^2 = .08$, $p < .05$. Planned simple effects contrasts revealed that the CR magnitude for faces in the 5 ($M = .05$, SD

[6] There was no significant difference in false alarms for those paired with old and new contexts ($p < .63$).

= .22) condition was significantly lower than the magnitude for faces in the 0 (*M* = .19, *SD* = .27) condition, $F(1, 49) = 9.24$, $MSE = .11$, $\eta_p^2 = .16$, $p < .01$. There were no other differences in CR magnitudes, although the difference between the 3 and the 0 pre-exposure conditions was marginal, $F(1, 49) = 3.43$, $p = .07$.

Relation between subjective association rating at encoding and CR magnitude

As in the previous experiments, I conducted Pearson correlations to examine the relation between CR magnitude and subjective 'match' ratings at encoding, in each pre-exposure condition. Correlations between mean subjective association ratings for faces and CR magnitude were not significant in the 0 ($r(50) = -.03$, $p = .82$), 1 ($r(50) = .27$, $p = .06$), 3 ($r(50) = .10$, $p = .48$), or 5 ($r(50) = .03$, $p = .86$) pre-exposure conditions.

2.4.3 Discussion

In Experiment 4, I replaced the 10 pre-exposures condition with a 5 condition in an effort to lower performance and reduce ceiling effects. Even after reducing the number of pre-exposures, I replicated the pattern of results observed in Experiment 3. Importantly, the CR magnitude was largest for the least familiar faces (0 pre-exposures) and smallest for the most familiar faces (5 pre-exposures). Additionally, the magnitude decreased incrementally as the degree of familiarity increased. In replicating the pattern of results and lowering accuracy rate performance, I was able to determine that the pattern could not be accounted for by ceiling effects. Instead, the pattern must be a result of the influence of number of pre-exposures on the CR effect. This pattern provides clear support for the outshining hypothesis in that the magnitude of the CR effect was completely eliminated for familiar faces.

Second, I again found no reliable correlations between subjective match ratings for targets to contexts, made during encoding, and magnitude of the CR benefit. This suggests that for faces, context effects are not influenced by how easily one could link, bind or match the target faces to the encoding context.

Comparing subjective association rating at encoding, across experiments

To investigate whether subjective 'match' ratings at encoding, in each pre-exposure condition, differed across experiments, I conducted a mixed 4 (Experiment: 1, 2, 3, 4) × 4 (Number of Pre-Exposures: 0, 1, 3, 10 (or 5× for Experiment 4)) ANOVA with Experiment as a between- subjects factor and Pre-Exposure as a within-subject factor. Results revealed a significant effect of Experiment ($F(3, 196) = 156.72$, $MSE = 1.22$, $\eta_p{}^2 = .71$, $p < .001$), whereby the mean subjective association rating for faces, both in Experiment 3 ($M = 3.99$) and 4 ($M = 3.94$), were significantly higher than for objects in Experiment 2 ($M = 2.45$; $p < .001$) or words in Experiment 1 ($M = 2.10$; $p < .001$). Association ratings for objects were also significantly higher than for words, $F(1, 196) = 9.74$, $MSE = .31$, $\eta_p{}^2 = .05$, $p < .01$.

There was also a main effect of the within-subject factor of Pre-Exposure even after applying Greenhouse-Geisser corrections, $F(2.72, 533.34) = 30.88$, $MSE = .16$, $\eta_p{}^2 = .14$, $p < .001$. Items pre-exposed 0 times ($M = 3.00$) had significantly lower association ratings than those pre-exposed 3 ($M = 3.26$; $F(1, 196) = 34.47$, $MSE = .40$, $\eta_p{}^2 = .15$, $p < .001$) and 10 ($M = 3.23$; $F(1, 196) = 41.93$, $MSE = .26$, $\eta_p{}^2 = .18$, $p < .001$) times; such a finding suggests that highly familiarized items may be more easily bound to an associated target, as suggested by Reder and colleagues (2013). Items pre-exposed 1 time ($M = 2.99$) also had significantly lower mean ratings than those pre-exposed 3 times ($F(1, 196) = 53.90$, $MSE = .28$, $\eta_p{}^2 = .22$, $p < .001$) and 10 ($F(1, 196) = 52.64$, $MSE = .23$, $\eta_p{}^2 = .21$, $p < .001$). Ratings for items pre-exposed 0 and 1 time did not differ.

Importantly, there was also a significant Experiment × Pre-Exposure interaction, $F(9, 588) = 11.54$, $MSE = .14$, $\eta_p^2 = .15$, $p < .001$. For the 0 pre-exposure condition, mean subjective association ratings were higher for faces (in Experiment 3) than for words in Experiment 1, ($F(1, 196) = 147.94$, $MSE = .48$, $\eta_p^2 = .43$, $p < .001$) and objects in Experiment 2, ($F(1, 196) = 102.24$, $MSE = .48$, $\eta_p^2 = .34$, $p < .001$). Faces in Experiment 4 also had higher mean ratings than both words ($F(1, 196) = 164.73$, $MSE = .48$, $\eta_p^2 = .46$, $p < .001$) and objects ($F(1, 196) = 117.54$, $MSE = .48$, $\eta_p^2 = .38$, $p < .001$). Similar patterns arose in the other pre-exposure conditions as well. In general, mean subjective association ratings were much higher for faces (in both Experiments 3 and 4) than for words or objects. The interaction can be accounted for by the fact that mean ratings did not differ for faces in Experiments 3 and 4 in the 0 times ($F(1, 196) = .87$, $p = .35$), 1 time ($F(1, 196) = .18$, $p = .67$), 3 times ($F(1, 196) = .31$, $p = .58$), or in the 5/10 times ($F(1, 196) = 2.59$, $p = .11$) pre-exposure conditions.

2.5 General Discussion

Across four experiments, I investigated whether familiarity with various types of target stimuli would influence context effects on recognition. Past work in the field has often used a wide variety of test stimuli, context types and familiarity manipulations. By maintaining a constant context type and familiarity manipulation, I could examine how stimulus type is affected by the CR effect. In my first two experiments, I used words and objects, respectively, as target stimuli. For the third and fourth experiments, I substituted unfamiliar faces for the word and object stimulus sets. Familiarity with target stimuli was manipulated using a pre-exposure phase, in which target stimuli were presented either 0, 1, 3, or 10 times in Experiments 1-3, and 0, 1, 3, or 5 times in Experiment 4. Recognition accuracy performance was assessed using an incidental test phase.

In the first experiment, I found that increasing familiarity, by way of number of pre-exposures, did not influence the magnitude of the benefit of context reinstatement on memory for words; similarly, large CR effects were observed regardless of pre-exposure condition. Along the same lines, for objects in the second experiment, though there was a significant interaction between pre-exposures and test context condition, I found significant CR effects in all pre-exposure conditions. As well, I found that increasing the number of pre-exposures from 1 to 3 to 10 did not further reduce the CR effect. The lack of impact of familiarity on word and object context effects is not overly surprising. There are several theoretical accounts which suggest that reinstating the encoding context during retrieval will always benefit memory, regardless of familiarity (CHARM: Eich, 1982; ICE: Murnane & Phelps, 1999; TODAM2: Murdock, 1997). I will discuss these models in greater depth in the next section.

As noted by other researchers, there is, however, mixed evidence for the effect of face familiarity on the CR effect. On one hand, some found evidence that the CR effect is enhanced for famous faces (Reder et al., 2013); on the other hand, there is evidence that the CR effect diminishes when faces become more familiar (Dalton, 1993). In Experiments 3 and 4, I investigated whether systematically increasing familiarity for faces would reduce the magnitude of the CR effect. In both experiments, I found that highly familiar faces received no benefit from context reinstatement. That is, CR magnitude was completely eliminated for faces that were pre-exposed 10 times in Experiment 3, and 5 times in Experiment 4. Specifically, accuracy for faces paired with the same contexts did not differ from accuracy for those paired with new contexts. Critically, the magnitude of the CR effect for the most familiar faces was significantly lower than the CR magnitude for the least familiar faces. These findings suggest that, for faces, the classic CR effect primarily applies to those that are unfamiliar or novel. Once a face has become familiar, context effects are reduced and other factors, such as signal strength, are more important in determining recognition performance.

2.5.1 Why are faces special?

The stark contrast in memory performance patterns for words and objects, in comparison to faces, suggests that faces are special. That is, familiarity seems to have a unique influence on how context interacts with, and benefits, memory for a face. One possible explanation is that, unlike for words and objects, faces have an inherent social component. Indeed, face recognition is an integral social ability (Hou & Liu, 2019) and can play a key role in forming and maintaining social relationships (Dalrymple et al., 2014; Fine, 2012; Yardley et al., 2008). As social constructs, it may be integral for face recognition to transcend context as people and families nomadically move from place to place. Words and objects, on the other hand, have arguably a very different purpose. Words, in particular, are for communication. In this way, the context surrounding a word will always drastically alter a given word's meaning. Our interpretation of the word 'glass' differs depending on the accompanying context (other words). If we say, "I would like to drink a *glass* of water", we most likely intend to convey that we are thirsty and want to drink water from a glass cup. Conversely, if we say something like, "I want to see the garden through this *glass* window", it conveys a very different meaning of the word. Here it conveys wanting to look outside of a window made out of glass. The purpose of a word is informed by the context as well. For example, the word *kick* is most often a verb or an action involving hitting an object with your foot (e.g., "kick the ball"). However, in other contexts, a kick can also be a noun, as in a sudden jolt (e.g., "the car started after giving it a kick"). As shown in these examples, the meaning of a word is *always* dependent on the context in which it was presented. Thus, a word (unlike a face) relies on its context to provide meaning and, consequently, memory for the word is always linked to the context in which it was presented.

Similarly, an object can hold many different meanings depending on the given context and situation. For example, a glass cup could be remembered as an object to aid drinking water, or as a container for dice in a game, depending on the context accompanying it. In contrast, faces must be recognizable across a variety of contexts, to facilitate recognition of threat and to maintain social relationships.

2.5.2 Theoretical Accounts

In general, theoretical accounts of the CR effect describe two potential influences of target familiarity. In my experiments, I found support for both of the contrasting hypotheses. The first, as described by global matching models (CHARM: Eich, 1982; TODAM2: Murdock, 1997) and more recent models, such as ICE (Murnane & Phelps, 1999), suggests that target-context pairings are encoded as an ensemble and, therefore, should always receive benefit from context reinstatement. The pattern of results that I found when manipulating familiarity for words and objects closely resembles this description. Importantly, my study was the first to show, in Experiment 3, that this pattern was not evident when the target is a face; such a finding is more in line with the outshining hypothesis, which argues that as target items become more familiar, context at retrieval becomes less beneficial (Smith & Vela, 2001). This pattern was replicated in Experiment 4.

My findings suggest that a refinement is needed in describing the effect of context reinstatement on memory. Context indeed benefits memory for words and objects, regardless of familiarity, but memory for target faces does not follow this pattern. I propose that because of their importance for social interactions, familiar faces may trump any role of context in modulating memorability.

2.5.3 Ease of linking target to context

Another potential pattern of results was proposed by Reder and colleagues (2013), who suggested that the familiar faces of celebrities can be more easily bound to new contexts as a result of having pre-existing representations. Contrary to other related work (Dalton, 1993), Reder et al. found that famous faces actually received a greater memorial benefit from context reinstatement at retrieval. In contrast to what Reder and colleagues (2013) found, recognition data of these studies suggest that familiar faces receive diminished, or no, benefit of context reinstatement. As well, I found no reliable correlation between magnitude of CR benefit and ratings of degree of 'match' between target and context, made during encoding. I acknowledge that my manipulation of face familiarity differed from the use of famous faces in the Reder et al. (2013) study. Celebrities and faces of famous people are often found in, and associated with, a large number of scenes and contexts. It may be easy to identify a famous person's face in new contexts because we have become accustomed to identifying them across a variety of situations. For example, it might be easy to identify actors such as Tom Hanks who have been in a large number of movies, spanning several genres. However, it may be more difficult to identify a celebrity's face when they are strongly associated with only one particular context, much like the butcher in his butcher shop (Mandler, 1980). In the future, researchers might be more likely to replicate the findings of Reder and colleagues (2013) if they examine the effect of pre-exposures on memory for faces encoded in a variety of scenes, as opposed to a faces associated with single contexts, as in the work presented here.

Interestingly, I also found that subjective ratings of association, or 'match' between target faces and their corresponding context scenes during the encoding phase were much higher than for words and objects with their corresponding contexts. This suggests that faces, in general, can be seen to plausibly 'match', or fit, with any number of context scenes. The same cannot be said for words and objects, for which ratings of association with contexts were lower (and more variable across pre-exposure condition). Thus, the relative importance of a context in informing memory for target faces, relative to target words or objects, may be reduced.

One difference between my study and the work of Reder and colleagues (2013) was that my manipulation of familiarity was performed using a pre-exposure manipulation. Reder and colleagues (2013) instead manipulated familiarity using famous, as opposed to non-famous, faces. By using two categories of stimuli, their manipulation allowed them to maintain their manipulation across encoding and retrieval phases. That is, lures used in their recognition test also fell under the categories of famous and non-famous. In my study, by contrast, all lures were unfamiliar to participants. It may therefore be important to consider how familiar, or famous, lures can also influence the CR effect. I note, however, that in my own unpublished study comparing context effects on famous and non-famous faces (in which lures were either famous or non-famous) we found no influence of fame on the CR effect. In fact, those results hinted at a reduced CR effect for famous faces. Future research should examine whether, and how, familiarity with lures influences the CR effect. Indeed, in the study by Reder and colleagues (2013), false alarm rate was much higher when lures depicted famous, compared to non-famous, faces. It is therefore possible that the level of familiarity of a target, relative to the lures presented during retrieval, may influence the CR effect differently than what I report here.

2.5.4 Limitations

In the pre-exposure phase, I had participants complete an orienting task to encourage processing of the target. A potential issue, however, is that each of the questions for this task differed depending on stimulus type. That is, for words, we asked participants to indicate whether the word contained the letter ‘a’, arguably a shallow level-of-processing question, whereas living/non-living and male/female distinctions were made for objects and faces, respectively. Not only did the stimulus types vary but potentially so did the extent to which items were encoded. As revealed by a large literature to date, greater depth of processing (Craik & Lockhart, 1972) can have a clear benefit to memory and an impact on the CR effect (Dougal & Rotello, 1999). While one might think that level of processing differences might be important, keep in mind that I found a similar pattern, of a sustained CR magnitude regardless of familiarity, when the target items were words and objects. Although I recognize the potential influence that levels-of-processing may have, in this particular instance, it seems context effects are still evident for familiar words and objects, but not for familiar faces.

Another factor to consider is whether I had sufficient power to detect a significant Pre-exposure by Test Context interaction in each experiment. Post-hoc power analysis of Experiment 1 indicates that, with an effect size of $\eta_p{}^2 = .03$, I had a power of .94, with the current sample size of 44. Sample sizes were even higher in Experiments 2, 3 and 4, making them sufficiently powered as well. Thus, I had roughly the same sample size and design across all four experiments, yet found different patterns of results only when faces were the target stimuli. Given the consistent manipulation and results of the post hoc power analysis, it is unlikely that the results are due to low power in Experiments 1 and 2, in which the CR effect was maintained in all pre-exposure conditions, relative to 3 and 4, in which the CR effect was reduced significantly for the highly familiar faces.

2.5.5 Conclusion

In four experiments, I examined whether familiarity with various stimulus types influenced the presence and magnitude of benefit of context reinstatement. Results indicated that familiarity has little to no impact on the presence and magnitude of the CR benefit, at least when the targets are words or objects. The CR benefit, however, was diminished for highly familiar faces. Results suggest that the CR benefit applies mainly when target stimuli are words or objects, regardless of item familiarity. For faces, the CR benefit applies only when faces are relatively novel; once familiar, other factors, such as target signal strength, guide memory performance.

Chapter 3: An Attentional Capture Account: Eye Tracking Experiment

Why might target familiarity impact the magnitude of the CR effect? A potential explanation may lie in an attentional capture account. Arguably, words require context to provide interpretation and meaning. That is, the definition of a word can be altered dramatically by its context. For example, the word *object* can be a noun as in the sentence “I picked up the object”. However, in other instances, it can be a verb with a very different meaning, such as in courtroom contexts when a lawyer shouts “I object!” In this instance, the word *object* not only forms two different parts of speech, but also the meaning of the word is very much changed by the context that it is used in. A face, on the other hand, can be processed and understood independent of its context. That is, relative to a word, a person’s face, or identity, will remain unchanged, regardless of the context. As such, processing of words may require attention to be deployed to the surrounding context to derive meaning. Conversely, faces would not.

In addition to replicating memory performance results observed in Chapter 2, I aimed to determine whether an attentional account could explain why the CR benefit is influenced by target face familiarity. Previous research has used eye tracking as an index of attention and has shown that eye gaze is an integral indicator for eventual successful face recognition (Althoff & Cohen, 1999; Henderson, Williams, & Falk, 2005; Peterson & Eckstein, 2012; Sterling et al., 2008) and that greater overall number (Olsen et al., 2016) and longer durations (Vakil, McDonald, Allen, Vardi-Shapiro, 2019) of fixations is associated with better memory. Of particular importance for the current study is whether familiarity impacts eye gaze patterns. Althoff and Cohen (1999) examined participant viewing patterns when presented with famous and non-famous faces in two blocks separated by two weeks. Results from eye tracking data revealed that, relative to non-famous face, participants spent more time gazing at the eyes of famous faces and less time towards the mouth. This pattern suggests that face familiarity may indeed have an influence on how a participant processes the target item during encoding. Heisz and Shore (2008) later replicated these results and further extended the findings by systematically increasing number of face exposures over the course of four days. With increasing exposures, participants made fewer fixations on the face during both recall and recognition tests. Crucially, the linear decline in fixation count was mirrored by a linear increase in both recall and recognition performance. This pattern suggests that fixation count may have a direct relation to later face memory. Previous research has also revealed that eye movements are more consistent when retrieval and encoding contexts match (Holm & Mäntylä, 2007), leading to an enhanced CR benefit. Therefore, it is reasonable to suggest that increasing face familiarity may indeed influence eye gaze behaviours, subsequently altering memory performance and the benefit of context effects. In this study, I used eye tracking as a measure of overt attention to determine whether eye gaze changes related to target familiarity can provide an explanation for any changes in context reinstatement magnitude.

In the current study, I manipulated familiarity using a similar pre-exposure paradigm as in the previous studies, whereby target items were presented to participants either 0, 1, 3, or 5 times. In this way, I could measure the influence of context reinstatement during retrieval, for target words and faces made systematically more familiar. As in the previous experiments from Chapter 2, I anticipated that increasing familiarity with words would not decrease the CR benefit. I also predicted that, as face stimuli became more familiar (i.e. the number of pre-exposures increased) the benefit of context reinstatement would decrease.

Importantly, I anticipated that eye tracking patterns would mirror these changes in memory performance. That is, as faces, in particular, become more familiar, eye gaze count and duration will decline, while recognition performance simultaneously increases. Some research has placed particular importance on gaze reinstatement for later memory (Wynn, Shen, & Ryan, 2019). This suggests that, if eye gaze patterns are different for familiar faces, then the way in which familiar faces are encoded may influence the nature of the ensemble or pairing that is formed, thus influencing later recognition. Therefore, eye gaze patterns may offer an explanation for the diminishing benefit of context reinstatement.

3.1 Method

Participants

Fifty-eight undergraduate students completed the study (40 females, *M* age = 19.75, *SD* = 1.80, *Range* = 17-25 years). Participants were enrolled in undergraduate Psychology classes and received course credit or token monetary remuneration for their participation. They each had normal, or corrected-to-normal vision.

Materials

Words. A set of 96 words were chosen from the Affective Norms for English Words (ANEW; Bradley & Lang, 1999). Selected words were concrete nouns.

Faces. A set of 96 faces were selected from the Center for Vital Longevity Face Database (Minear & Park, 2004). Faces included those of young to middle-aged adults from an assortment of ethnicities to match the heterogeneity of the sampled student population. Photographs included head and shoulders only and contained no distinguishing features (e.g. facial accessories, glasses, sunglasses, hats). Faces were presented in a frontal view, in colour, and on a white background.

Scenes. Context stimuli consisted of scene photographs (half indoor/half outdoor), presented in colour, collected through an Internet search of publicly accessible images. Scenes included images of outdoor (e.g. beach, park, basketball court) and indoor scenery (e.g., living room, restaurant) and did not include visible faces or words.

Design and Procedure

Eye tracker. Prior to each word and face blocks, participant eye gaze was calibrated to an eye tracker. An Eyelink 1000 desk-mounted remote system (SR Research, http://sr-research.com), sampling at 1000 Hz, was used to measure eye position. At the beginning of each block the participant's dominant eye (as determined by the Miles test; Miles, 1930) was calibrated using a nine-point automatic calibration sequence. The eye tracker was calibrated prior to the first word or face block (depending on which block the participants first completed) and was re-calibrated between word and face blocks. Participant eye gaze was calibrated using a 9-point calibration test, followed by a validation phase. See Figure 3 for procedure.

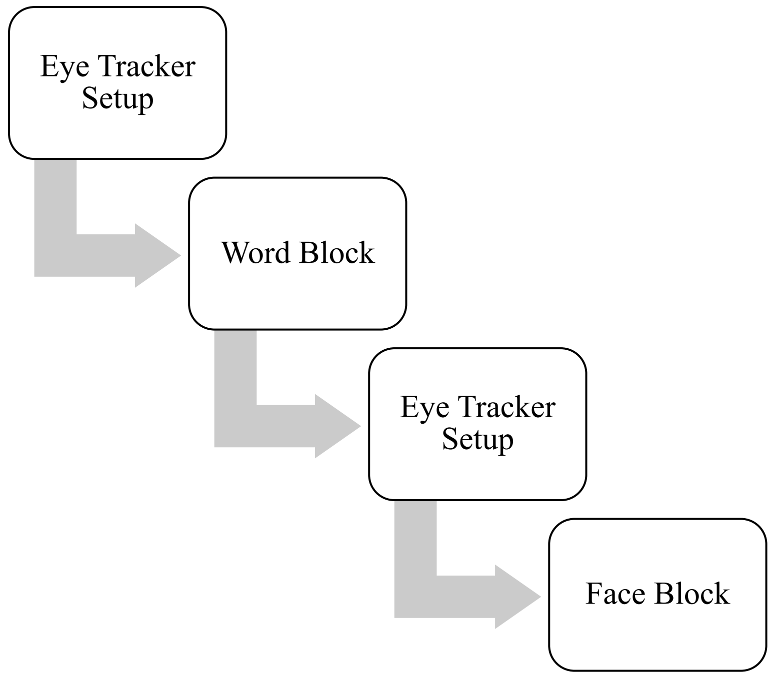

Figure 3. Study procedure. Participants first calibrated gaze to an eye tracker, followed by either the word or face block, counterbalanced, (word block sample is shown above). Once the block was completed, participants recalibrated eye gaze before completing the remaining word or face block.

Word and face blocks. The sets of 96 words and 96 context photographs were randomly paired. The set of 96 words and 96 context photos were randomly paired. The pairs were then randomly divided in half to create two study lists (List A and List B), each consisting of 48 words and 48 context scenes (half indoor/half outdoor) pairs. List A was presented to half of the participants during the study phase and List B was presented to the second half of participants. The set of faces were subdivided using the same procedure as the words.

Each of the word and face blocks were subdivided into three phases. The first phase, a pre-exposure phase, was administered to familiarize participants with the target items (either words in the word block, or faces in the face block). Thirty-six of the 48 items from the predetermined list (A or B) were divided into three group conditions. Amongst the 36 study items, 12 were pre-exposed once, 12 were pre-exposed 3 times and 12 were pre-exposed 5 times.

During each pre-exposure trial, the target item was presented for 3500 msec, followed by a fixation cross for 500 msec. Words (or faces) were presented alone on a white background. Items were presented in a randomized order. In Word blocks, words were presented in black in 32-point Arial font, and for Face blocks, faces were presented in colour. For each of the word and face blocks, participants were asked to make a binary decision. For words, participants were asked to indicate if the word contained the letter 'a' or not, and for faces, whether it was of a male or female. Participants made responses, for each block, by pressing either the "v" or "b" key on a standard QWERTY keyboard.

The second phase consisted of an incidental encoding phase for the Word (or for the Face block, in counterbalanced order). Participants were presented with 48 target word-context pairs from the predetermined list. Each pair was presented on the monitor for 2250 msec in a randomized order. Following the presentation of the pair, a subsequent screen displaying a Likert scale appeared. Participants were instructed to make a rating on the scale from 1 to 6 indicating how likely the two stimuli belonged together. The scale remained on screen until the participant made a response. Participants were instructed to use the top row of numbers to make their response. A sample of each pair type (word-context or face-context) is shown in Figure 1A and 1C in Chapter 2.

The third phase consisted of a surprise, incidental recognition test. Target words (or faces) from the previous phase were either presented alongside the same context as during the encoding phase, or alongside a novel context that had not been shown previously in either the face or word blocks. That is, half (24) of the encoding pairs remained intact (*same* context) and half were presented with a *new* context. The remaining half of the context scenes from the rearranged pairs were paired alongside foil words (or foil faces). In addition, there were six pairs consisting of both novel words and novel context scenes. There were therefore 4 trial pair type on the recognition test: 1) old target word (or face) + same context, 2) old target word (or face) + new context, 3) foil word (or face) + old context, 4) foil word (or face) + new context.

During each trial of the recognition test, participants were instructed to indicate if the target item (either a word or face in the word or face block, respectively) had been presented in the previous phase (encoding phase) or not by pressing "1" to indicate they had seen the target previously or "3" if they had not seen the target before, on a standard QWERTY keyboard. Target pairs were presented on screen for 5000 msec. In the event that a key was not pressed within this allowed trial time, it was scored as missed responses, and the subsequent trial would ensue.

Study lists for words and for faces were counterbalanced across participants. That is, for each, List A was used as the target study list and List B was used as the foil list for half of the participants, and this was counterbalanced across participants. For both study and test, the word (or face) was always presented on the left and context scene on the right (see Figure 1). Importantly, the word and face block order was also counterbalanced across participants.

3.2 Results

3.2.1 Memory Accuracy for Targets

Memory performance for both words and faces was calculated for each participant, subdivided by Test Context (same, new) and Pre-Exposure (0, 1, 3, 5) conditions. Hit rate was calculated as the number of correct hits for words and faces in each condition divided by 6 (the total possible number of hits in each category). False alarm rate was calculated as the total number of false alarms, divided by the 24 when foils were paired with old context scenes, or 6 for foils paired with novel contexts. Accuracy rates for faces and words were calculated as the difference between hit rates and false alarm rates within each specified category. Significant interactions were further analyzed by contrasting accuracy rate means for each pre-exposure condition.

Data were excluded for 4 participants who scored accuracy rates of 0, in either the word or face recognition test. In total, data from 54 participants were included in the accuracy rate analysis. Accuracy rates were analyzed using a repeated-measures ANOVA with within-subjects factors of Stimulus Type, Test Context and Pre-Exposure condition. I also included Order (word block first followed by faces, or vice versa) as a between-subjects factor. Accuracy rate data for both faces and words are shown in Table 3.

Accuracy rate was analyzed using a 4 (Pre-Exposure condition: 0, 1, 3, 5) x 2 (Test Context condition: same, new) x 2 (Study Stimulus: face, word) x 2 (Block Order: face block first, word block first) repeated-measures ANOVA. Results revealed that block order did not affect any main effects or interactions (all *p's* > .55). Therefore, I conducted a 4 (Pre-Exposure condition: 0, 1, 3, 5) x 2 (Test Context condition: same, new) x 2 (Study Stimulus: face, word) ANOVA, and excluded study order from further analyses. Results revealed no significant 3-way interaction even after applying a Greenhouse-Geisser correction, $F(2.63, 139.63) = 2.10$, $MSE = 0.02$, $\eta_p^2 = .04$, $p = .11$[7].

Given my a priori hypotheses, I conducted two separate 2 (Test Context) x 4 (Pre-Exposure condition) ANOVAs for faces and words separately.

[7] Using G*Power 3 (Faul et al., 2007), we determined that to detect a significant difference with a small effect size ($f = .1$) at an alpha of .05, we would have needed 88 participants. Therefore, though trending in the hypothesized direction, we may have been underpowered to find the significant interaction we had anticipated.

Table 3

Hit rate, false alarm rate and accuracy rate for each stimulus type, pre-exposure condition, and recognition test context (standard deviation in parentheses).

	Context at Test	Hit Rate (SD)	ACC (SD)	Hit Rate (SD)	ACC (SD)	Hit Rate (SD)	ACC (SD)	Hit Rate (SD)	ACC (SD)	FA Rate (SD)
Words		0 Exposures		1 Exposure		3 Exposures		5 Exposures		
	Same	.88 (.26)	.80 (.33)	.85 (.26)	.77 (.34)	.90 (.25)	.82 (.32)	.89 (.26)	.81 (.33)	.08 (.15)
	New	.71 (.27)	.64 (.31)	.69 (.34)	.61 (.36)	.77 (.30)	.70 (.32)	.77 (.30)	.70 (.33)	.07 (.17)
Faces		0 Exposures		1 Exposure		3 Exposures		5 Exposures		
	Same	.67 (.25)	.58 (.24)	.83 (.20)	.74 (.21)	.93 (.16)	.83 (.17)	.96 (.15)	.86 (.18)	.10 (.10)
	New	.44 (.23)	.34 (.27)	.67 (.24)	.57 (.26)	.85 (.23)	.75 (.25)	.88 (.26)	.78 (.28)	.10 (.16)

Note: ACC = accuracy rate, FA = false alarm, SD = standard deviation

Faces

Results of the 2 (Test Context: same, new) x 4 (Pre-Exposure Condition: 0, 1, 3, 5) repeated-measures ANOVA revealed a significant main effect of Test Context, $F(1, 53) = 30.04$, $MSE = 0.07$, $\eta_p{}^2 = .36$, $p < .001$. As predicted, memory performance was enhanced when faces were paired with same ($M = .75$), as compared to new ($M = .61$), context scenes. Greenhouse-Geisser corrections were applied to the following analyses to correct for violations of the assumption of sphericity. Even after applying corrections, there was also a main effect of Pre-Exposure condition, $F(2.22, 117.74) = 126.67$, $MSE = 0.03$, $\eta_p{}^2 = .71$, $p < .001$. Accuracy rates were significantly lower for faces pre-exposed 0 times compared to those pre-exposed 1 time ($F(1, 53) = 60.35$, $MSE = 0.07$, $\eta_p{}^2 = .53$, $p < .001$), 3 times ($F(1, 53) = 186.63$, $MSE = 0.07$, $\eta_p{}^2 = .78$, $p < .001$) or 5 times ($F(1, 53) = 214.71$, $MSE = 0.07$, $\eta_p{}^2 = .80$, $p < .001$). Accuracy rates were also significantly higher for faces pre-exposed 3 times compared to 1 time ($F(1, 53) = 76.93$, $MSE = 0.03$, $\eta_p{}^2 = .59$, $p < .001$). However, accuracy rate for faces pre-exposed 3 and 5 times did not differ ($p < .05$). All contrasts were still significantly different even after Bonferroni corrections were applied.

Critically, there was also a significant Test Context x Pre-Exposure condition interaction ($F(2.26, 119.65) = 6.91$, $MSE = 0.03$, $\eta_p{}^2 = .12$, $p < .001$). A priori planned paired sample t-tests revealed that the difference in accuracy rates between faces paired with the same and new contexts was the largest for faces pre-exposed 0 times, ($M_{\text{same}} = .58$, $M_{\text{new}} = .34$; $t(53) = 4.81$, $p < .001$). Though all differences were significant, the differentiation in accuracy rate gradually diminished as the number of pre-exposures increased. That is, the benefit of context reduced as faces became more familiar. Specifically, the context benefit for faces pre-exposed 1 time ($M_{\text{same}} = .74$, $M_{\text{new}} = .57$; $t(53) = 4.77$, $p < .001$) was slightly larger than for faces pre-exposed 0 times, and smaller than the difference for faces in the 3 pre-exposure (M_{same} = .83, M_{new} = .75; $t(53)$ = 3.22, $p < .01$) condition. Importantly, the context benefit was the smallest for faces in the 5 pre-exposure condition (M_{same} = .86, M_{new} = .78; $t(53) = 2.58$, $p < .05$).

Words

Results of a 2 (Test Context: same, new) x 4 (Pre-Exposure Condition: 0, 1, 3, 5) repeated-measures ANOVA revealed a significant main effect of Test Context, $F(1, 53) = 17.07$, $MSE = 0.13$, $\eta_p{}^2 = .24$, $p < .001$. As with faces, accuracy rate was higher when words were paired with the same ($M = .80$), compared to new ($M = .66$), context as at encoding. There was also a main effect of Pre-Exposure Condition, $F(3, 159) = 6.15$, $MSE = 0.02$, $\eta_p{}^2 = .10$, $p < .001$). Simple effects contrasts revealed words pre-exposed 0 times were recognized significantly less than those in the 3 pre-exposures ($F(1, 53) = 6.67$, $MSE = 0.03$, $\eta_p{}^2 = .11$, $p < .05$) condition, but did not differ from words in the 1 and 5 times conditions (p's < .05). Accuracy rate for words pre-exposed one time was significantly lower than those pre-exposed 3 ($F(1, 53) = 13.88$, MSE = 0.03, $\eta_p{}^2 = .21$, $p < .001$) and 5 ($F(1, 53) = 8.44$, $MSE = 0.05$, $\eta_p{}^2 = .14$, $p < .01$) times. There was no difference between words pre-exposed 3 and 5 times ($p = .71$). Critically, and in contrast to the results with faces, the interaction between Test Context and Pre-Exposure condition was not significant ($p = .24$).

3.2.2 Eye Tracking Data Analysis

Eye tracking data from the initiation of the encoding phase until the target-context pair disappeared were recorded. This strategy ensured exactly 2.25 seconds of eye tracking recording was measured during each trial during encoding. For eye tracking data, rectangular regions of interest were created, spanning a box containing the location where words, faces and context scenes were presented, were created. Data from 22 participants were excluded due to technology malfunction on two separate occasions. In total, data from 32 participants were included in the eye tracking analysis[8].

As in (Althoff & Cohen, 1999; Sterling et al., 2008), I included measurements of dwell time and fixation count. For each of these variables, I calculated mean proportions of time spent directing gaze towards either the target (face, word), and the context scene. Proportions were then subdivided into pre-exposure conditions (0, 1, 3, 5) and analyzed using a repeated-measures ANOVA to determine whether attention varied depending on pre-exposure condition.

Fixation Count

Mean fixation count proportions are displayed in Figure 4. Fixation counts were defined as when gaze remained directed at a single location for at least 80 msec. I first analyzed data in a 2 (Eye Gaze Location: target, context) x 4 (Pre-Exposure condition: 0, 1, 3, 5) x 2 (Stimulus Type: faces, words) x 2 (Block Order: face block first, word block first) ANOVA. Results of the ANOVA revealed no main effect nor any interactions (all *p's* > .23) with Order. I therefore conducted a 3-way ANOVA excluding block order as a factor.

Results of this ANOVA revealed no main effects of Eye Gaze Location, Pre-Exposure condition or Stimulus Type. There was, however, a significant Eye Gaze Location x Pre-Exposure interaction, $F(3, 93) = 7.20$, $MSE = 0.003$, $\eta_p^2 = .19$, $p < .001$. This interaction could be accounted for by a fluctuation in the eye gaze location depending on the number of pre-

[8] Patterns of accuracy rate results when analyzing data from 32 participants were similar to those when analyzing all 54 participants. There was no significant 3-way interaction ($F(3, 93) = 1.23$, $p = .30$). Though the face accuracy Test Context x Pre-Exposure condition interaction was not significant ($F(1.91, 59.24) = 2.70$, $p = .08$), there was a similar trend in the data. That is, the CR magnitude decreased as number of pre-exposures increased. Moreover, for words, there was again no significant interaction between Test Context and Pre-Exposure condition ($F(3, 93) = 1.47$, $p = .23$), whereby CR magnitude remained similar, regardless of number of pre-exposures.

exposures. That is, the ratio in target to context fixation counts was very similar in the 0 (M_{target} = .48, M_{context} = .50) and 3 (M_{target} = .48, M_{context} = .49) pre-exposures conditions, and much more variable in the 1 (M_{target} = .46, M_{context} = .51) and 5 (M_{target} = .46, M_{context} = .52) times conditions.

There was also a significant Eye Gaze Location x Stimulus Type interaction, $F(1, 31) = 22.08$, $MSE = 0.05$, $\eta_p^{\,2} = .42$, $p < .001$. This interaction could be accounted for by two diverging eye gaze patterns depending on stimulus type. For faces, a higher number of fixations were directed towards the target (M_{target} = .52, M_{context} = .46). In stark contrast, the exact opposite pattern was seen with words, whereby a higher ratio of fixation counts were directed towards the context than the target words (M_{target} = .42, M_{context} = .55). The Pre-Exposure condition x Stimulus Type interaction was not significant.

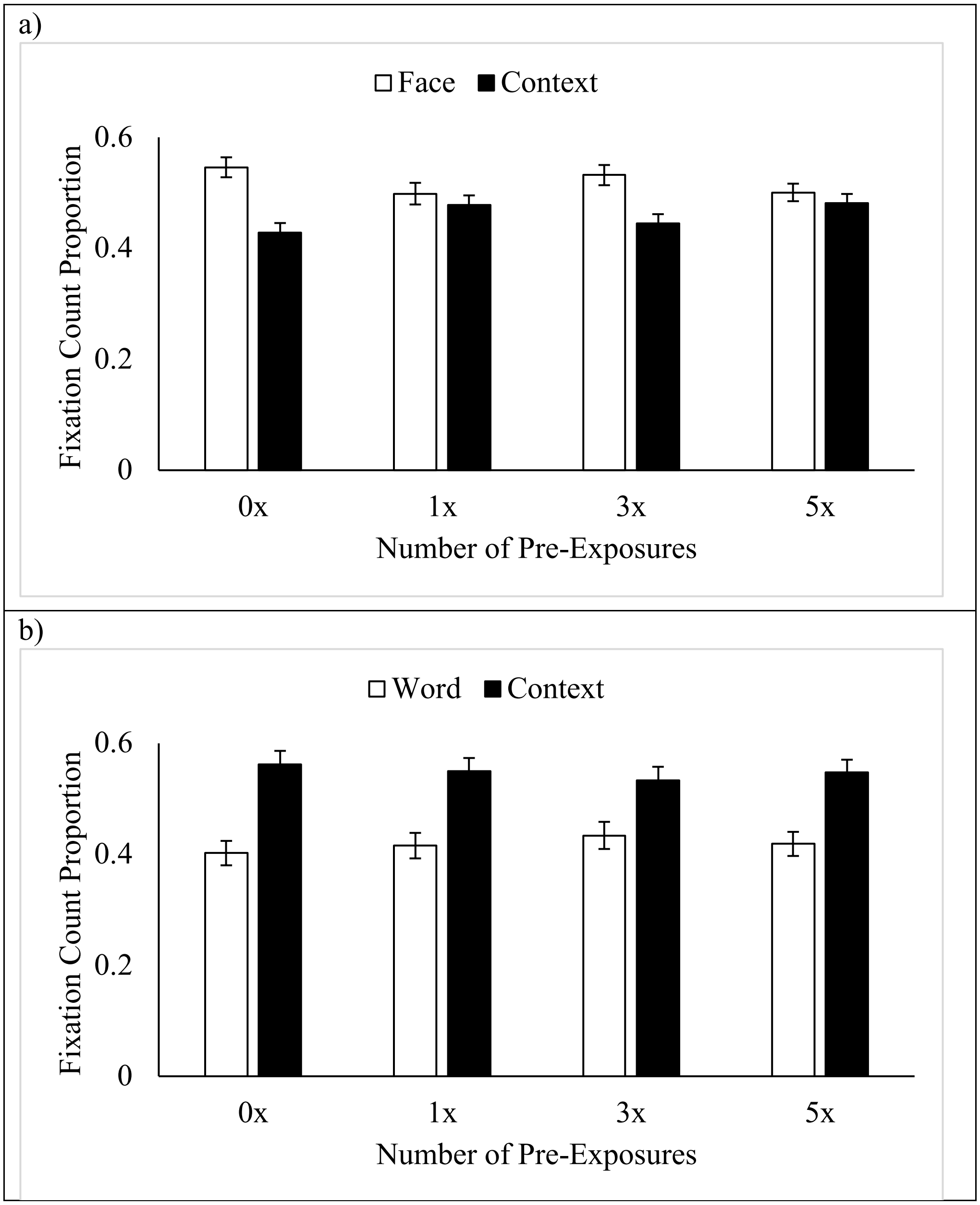

Figure 4. Mean proportion of fixation counts directed towards the face, in (a) face blocks, or words, in (b) words blocks, or accompanying context scenes.

There was also a significant 3-way interaction, $F(3, 93) = 8.77$, $MSE = 0.003$, $\eta_p{}^2 = .22$, $p < .001$. I conducted two separate 2-way ANOVAs (Eye Gaze Location: target, context; Pre-Exposure condition: 0, 1, 3, 5) for the face and word blocks to further break down the interaction. The 3-way interaction could be accounted for by two distinct patterns for faces and for words. For faces, results of the ANOVA revealed a 2-way interaction ($F(3, 93) = 12.10$, $MSE = 0.003$, $\eta_p{}^2 = .28$, $p < .001$), whereby regardless of pre-exposure condition, the number of fixations directed towards the target face ($M_0 = .55$, $M_1 = .50$, $M_3 = .53$, $M_5 = .50$) was always higher than the number of fixations directed towards the context ($M_0 = .43$, $M_1 = .48$, $M_3 = .45$, $M_5 = .48$). The number of pre-exposures also influenced fixation count, such that the proportions were much more similar in the 1 ($t(31) = .64$, $p = .53$) and 5 ($t(31) = .65$, $p = .52$) pre-exposures conditions, but significantly differed for the 0 ($t(31) = 3.88$, $p < .001$) and 3 ($t(31) = 2.66$, $p < .05$) pre-exposures faces. When words were the target, the repeated-measures ANOVA revealed a significant 2-way interaction $F(3, 93) = 3.39$, $MSE = 0.003$, $\eta_p{}^2 = .10$, $p < .05$. In contrast to faces, however, there was a significantly higher number of fixations directed towards the context than the target word in each pre-exposure condition (0 pre-exposures: $M_{\text{context}} = .56$, $M_{\text{target}} = .40$, $t(31) = 4.21$, $p < .001$; 1 pre-exposure: $M_1 = .55$, $M_1 = .42$, $t(31) = 3.46$, $p < .01$; $M_3 = .53$, $M_3 = .43$, $t(31) = 2.37$, $p < .05$; $M_5 = .55$, $M_5 = .42$, $t(31) = 3.62$, $p < .01$). Moreover, as opposed to the faces, whereby a fluctuation was observed, number of pre-exposures did not impact words in the same way. Instead, regardless of pre-exposure number, the proportion of fixation counts remained fairly consistent.

Dwell Time

Mean dwell time proportions are displayed in Figure 5. As with fixation count, I first analyzed dwell time data in a 2 (Eye Gaze Location: target, context) x 4 (Pre-Exposure condition: 0, 1, 3, 5) x 2 (Stimulus Type: faces, words) x 2 (Block Order: face block first, word block first) ANOVA. There was no main effect of Order, though the Block Order x Pre-Exposure condition interaction was significant, $F(3, 90) = 2.77$, $MSE < .001$, $\eta_p^2 = .09$, $p < .05$. No other interactions with Order were significant (ps > .05), therefore, I present the next set of analyses without Block Order as a factor.

There were no significant main effects. There was, however, a significant Eye Gaze Location x Stimulus Type interaction, $F(1, 32) = 63.67$, $MSE = .03$, $\eta_p^2 = .67$, $p < .001$. This interaction was explained by two distinct patterns. Specifically, when faces were the study target, a higher ratio of dwell time was dedicated towards the target than the context ($M_{\text{target}} = .57$, $M_{\text{context}} = .41$). The opposite pattern was observed when words were the study targets ($M_{\text{target}} = .43$, $M_{\text{context}} = .54$). There were no other significant interactions. The 3-way interaction was also non-significant, $F(3, 93) = .41$, $MSE = .07$, $\eta_p^2 = .01$, $p = .75$.

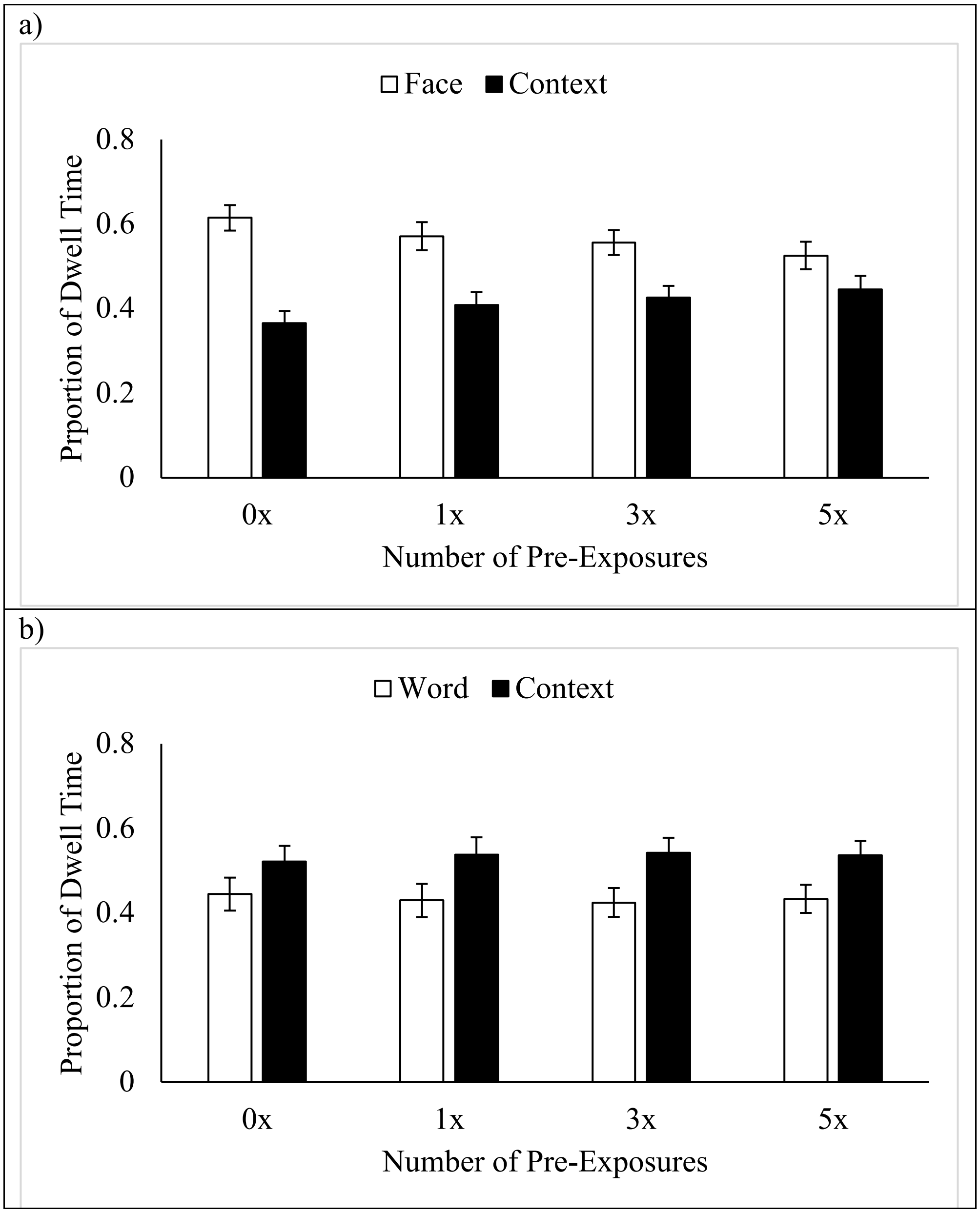

Figure 5. Mean proportion of dwell time directed towards the face, in (a) face blocks, or words, in (b) words blocks, or accompanying context scenes.

3.3 Discussion

In this study, I investigated whether familiarity with target faces and words influenced the benefit of context reinstatement to memory. I manipulated familiarity by pre-exposing targets 0, 1, 3, and 5 times. In addition, I measured participants eye gaze patterns using an eye tracker as an indirect measure of attention.

Results from experiments in Chapter 1 revealed that the context reinstatement effect was nearly completely diminished when faces became very familiar. This reduction, however, was not observed when words were used as the test stimuli. Accuracy rate results of this study demonstrated a similar pattern to that of my previous work. That is, in this study, though there was always a significant benefit of context in all pre-exposure conditions, the magnitude of the CR effect for faces declined gradually as the number of pre-exposures to target faces increased. The CR magnitude for words, however, was unaffected by pre-exposure condition and remained consistently large across all pre-exposure conditions.

This is a key finding as it provides evidence that, for faces, the reliance on context to determine memory decisions declines with increasing familiarity. Faces that are very familiar benefit less from context reinstatement, whereas relatively unfamiliar faces receive a comparatively much larger benefit. On the other hand, when words are the targets, the CR benefit is maintained regardless of how many times they were pre-exposed. Therefore, the CR effect seems to consistently apply to words, but applies on to faces when they are novel, rather than familiar. Though I did find that CR magnitude declined as familiarity increased, there was still a significant benefit of context reinstatement for faces in the 5 pre-exposure condition. One reason I may have been unable to detect a 3-way (Test Context, Stimulus Type, and Pre-Exposure condition) interaction in accuracy rate performance was because the study may have been underpowered to detect such an interaction. Using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), I determined that to detect a significant difference with a small effect size (f = .1) at an alpha of .05, I would have needed 88 participants. Therefore, though trending in the hypothesized direction, I may have been underpowered to find the significant interaction I had anticipated.

In addition to recognition test performance, I also recorded and analyzed proportion of fixation counts and dwell time directed towards either the target (faces or words in their respective blocks) or context. Using these two metrics, I found two converging patterns of results. In both eye tracking metrics, fixation count and dwell time, I found that a larger portion of these was directed towards the face than the associated context scene. Critically, the exact opposite was observed when words were used as the target, as a higher proportion of both eye gaze metrics were directed towards the context, instead of the word.

This preference to direct gaze to the context over a paired word may explain the pattern demonstrated in the recognition test. If a higher number of fixations and longer fixation duration is consistently devoted to the context at encoding, perhaps then, on the later recognition test, greater reliance is placed on the presence of the context scene in guiding memory decisions. In line with this suggestion, I observed a large CR benefit, regardless of familiarity of the target word. The exact opposite pattern was observed for faces, as these received the larger share of gaze during encoding, relative to the context. As these eye tracking measures favoured the face rather than context during encoding, it is possible that later memory for the face became less dependent on contextual cues.

In this study, I investigated the influence of familiarity on the CR effect for faces and words. In line with the outshining hypothesis, results indicated a reduction in CR effect for familiar faces. However, though the CR effect declined, there was still a small and significant benefit of CR for familiar faces. For words, a healthy-sized CR boost was observed, regardless of the number of pre-exposures. This consistent benefit of context, for target words, is suggested by global matching models. This experiment therefore offers support for both the outshining hypothesis (Smith & Vela, 2001) and the global matching models (TODAM2, Murdock, 1997; ICE model, Murnane & Phelps, 1999). In addition to memory performance, I also tracked eye gaze as an indirect measure of attention. Results revealed that participants tended to look longer at faces than their associated contexts during encoding, whereas the opposite was true for words. This pattern suggests that, when studying words, a relatively larger proportion of attention was always dedicated to processing peripheral cues and supporting details. Words therefore, may always be reliant on the context to provide meaning and aid memory, whereas faces are not. Instead, when faces are familiar, the benefit of CR gradually declines, potentially to the extent that it is no longer beneficial or required to guide memory decisions for faces.

Chapter 4: Who You Are and Where We Meet: Social Anxiety Influences CR Benefit

To this point, I have focused my PhD experiments on qualities of the target that influence the CR effect. Specifically, I maintained a consistent context and familiarity manipulation, while changing the target stimulus type, to determine the impact of stimulus type on the magnitude of CR effect. Important to consider is the other side of the target-context pairing. That is, not only can changing the target stimulus type impact the CR effect, but variations of the context could also be impactful. In this Chapter, I will discuss research I conducted and submitted with my colleague, Ryan Yeung (Yeung, Lee, & Fernandes, submitted), where we investigated how the social-anxiety provoking nature of a context scene could influence the benefit of the CR effect.

In one study by Krafka and Penrod (1985), store clerks were tested on their ability to identify a customer who had visited their store previously. By mentally reinstating the context (i.e., attempting to reconstruct the customer's face and the events of the customer's transaction) and having contextual cues physically presented to them (i.e., a piece of non-picture identification from the customer), accuracy of identifying the customer from a lineup was significantly enhanced. Similarly, Hammond, Wagstaff, and Cole (2006) asked participants to mentally reinstate the encoding context (i.e., trying to reconstruct the surrounding environment and one's own thoughts/feelings at the time) of a video of an armed robbery, without providing them any physical contextual cues. Results suggested that mental CR alone was sufficient to enhance memory for details including what the robber looked like. Lending further support, Wong and Read (2011) found that participants were more likely to correctly identify the culprit of a crime video when participants both encoded and retrieved in the same room. Given the diverse contexts that seem to provide a CR benefit for face memory, it would therefore seem reasonable to expect a CR effect regardless of characteristics of the context.

However, various factors related to the target items and their contexts do seem to systematically influence the CR effect. Specifically, prior work has identified that the manner in which items and contexts are presented relative to each other can significantly alter the CR effect. For example, Dalton (1993) demonstrated that failing to reinstate global contexts (e.g., context elements encoded to many items) impaired recognition of unfamiliar faces, whereas failing to reinstate local contexts (e.g., context elements encoded uniquely to one or few items) impair recognition of faces in general. Moreover, non-incidental contexts (e.g., contexts obviously related to the target memoranda) have been shown to provide an enhanced CR benefit relative to incidental contexts (e.g., contexts not obviously related to the target memoranda; Smith & Vela, 2001). Additionally, work by Baddeley (1982) suggests that reinstating an interactive context (e.g., when the context modifies the meaning of a target word, as in the pair *cold-ground*) enhances recognition, while reinstatement of independent contexts (e.g., when the context and a target word are encoded individually) provides no such CR benefit.

More recently, there has been interest in specific features of the context itself, rather than how the item and context are presented in relation to each other. In one such instance, Skinner and Fernandes (2010) found that displaying an intact face as a context item provided a greater CR benefit to target word memory compared to a context image of a face with its pixels scrambled. As such, it would seem that CR fails to boost memory when the context lacks semantic meaning. Given this evidence that features of the specific context being reinstated can modulate the CR effect, I investigated how the anxiety-provoking nature of a scene may affect its ability to provide a CR benefit.

Although there is little work directly assessing anxiety-provoking contexts, there is some evidence that high arousal (a potential component of anxious feelings), as provoked by a scene, may reduce the CR effect. For example, Brown (2003) had participants view a slideshow of an encounter between a cyclist and a pick-up truck, which was either neutral (the cyclist passing the truck unharmed) or emotionally arousing (the cyclist lying on the pavement in front of the truck, bleeding). Reinstating the arousing context provided no memory benefit to target information. In contrast, the classic CR effect was found when reinstating the neutral context.

In a similar vein, negative valence (another potential component of anxious feelings) in scenes has been shown to reverse the CR effect, such that reinstated contexts impair memory. For example, the reinstatement of emotionally negative contexts impairs memory for faces. In a study by Rainis (2001), faces were paired with either a negative, positive, or neutral background scene at encoding. Participants completed a recognition test six days later, during which faces were paired with either (1) the exact same context as encoding, (2) a context containing the same subject matter as encoding, or (3) a completely novel context scene. Results showed a deficit in memory for faces when these were paired with negative contexts, relative to positive or neutral contexts. This impairment was found even when the negative context was kept identical from encoding to retrieval (i.e., reinstated).

Overall, evidence supports the idea that emotionally arousing or negative contexts influence the CR effect. However, what mechanisms might drive these observed effects on CR? Potential explanations may stem from current models that suggest that the binding between an item and its context is weakened by strong emotionality, such that the ability for a context to support memory for its paired item would be reduced. According to one account, emotional arousal leads to enhanced processing of high-priority information, but impairs processing of low-priority information (Mather & Sutherland, 2011). In terms of memory performance, this arousal-biased competition model proposes that the shift in prioritization leads to impaired binding between an item and its context (relative to within-item featural binding) in paradigms where item-context binding is not the participant's explicit goal. In another framework, negative emotion also disrupts item-context binding through the downregulation of hippocampal activity, which is critical for forming associative memories (Bisby & Burgess, 2017). It would therefore seem that anxiety-provoking contexts, due to their tendency to promote high arousal and/or negative valence, present an interesting opportunity to test these models' predictions in the novel domain of CR.

In Experiment 6 and 7, I investigated whether anxiety-provoking context scenes would influence the magnitude of the CR benefit to memory for target faces, given that anxious feelings could induce both high arousal and negative valence. Participants incidentally encoded target faces paired with scenes that were either low anxiety-provoking or high anxiety-provoking (as confirmed by a validation study in naïve participants). I also systematically manipulated the presence of faces embedded within the context scene. At retrieval, participants performed a recognition test for old, target faces (intermixed with novel, distractor faces) that were paired with either the same or a new context scene. A key feature of my study was the validation of the context scenes to confirm anxiety-provoking quality. Unlike previous work that used fearful or highly arousing situations to investigate the CR effect (Brown, 2003; Hammond, Wagstaff, & Cole, 2006; Krafka & Penrod, 1985; Wong & Read, 2011), I confirmed that the scenes were anxiety-provoking to a naïve sample recruited from the same participant pool. Specifically, I adapted items from the Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987) to measure the degree to which the context scenes provoked anxiety. Further, participants also reported their current feelings of emotionality in response to viewing the scene, as indexed by the valence and arousal scales of the Self-Assessment Manikin (SAM; Bradley & Lang, 1994).

I hypothesized that highly anxiety-provoking context scenes would provide a reduced CR benefit to face memory. Potentially, embedded faces could be viewed as additional opportunities for social interaction and, in turn, provoke further state social anxiety. As a consequence, I predicted that the presence of such faces would further reduce the CR effect.

Finally, I measured memory for the context scenes themselves. Previous work has shown that contexts reinstated at test are better recognized (Hanczakowski, Zawadzka, & Macken, 2015). As such, I included a recognition test for the context scenes presented alone, following the recognition test for the target faces. I predicted that scenes used in the face recognition test as reinstated contexts would be better remembered. In addition, I anticipated a memory bias for high, in contrast to low, anxiety-provoking scenes.

4.1 Experiment 6

4.1.1 Method

Participants

Fifty undergraduate students enrolled in psychology courses completed the current experiment for partial course credit. Participation was restricted to individuals who reported having lived in Canada and/or the United States for at least 10 years, in order to ensure that all participants had prior experience with Caucasian faces. One participant was removed from all analyses due to a 0% accuracy score (hit rate minus false alarm rate) in the face recognition phase, leaving 49 participants in the final sample. Of the remaining participants, 73% were women, 27% were men, and the mean age was 21.1 (*SD* = 6.3).

Social Phobia Inventory

Due to my use of social anxiety-provoking contexts as stimuli, I also measured participants' trait level of social anxiety in order to assess if a general propensity to experiencing social anxiety might be related to memory performance in Experiments 1 and 2. The Social Phobia Inventory (SPIN; Connor et al., 2000) consists of 17 self-report items that evaluate fear, avoidance, and physiological discomfort related to various social scenarios. Participants indicated on a scale from 0 (*not at all*) to 4 (*extremely*) the extent to which a series of problems (e.g., "Being embarrassed or looking stupid are among my worst fears") tended to bother them in a typical week. Although the established cut-off score of 19 is thought to have some ability to distinguish between individuals with and without social anxiety disorder, formal diagnoses of social anxiety disorder or any other psychopathology were not queried, nor confirmed by a clinician. The SPIN has been found to have good test-retest reliability, internal consistency, convergent and divergent validity, and construct validity (Connor et al., 2000). In the current study, mean SPIN scores were 26.3 (SD = 14.2) for Experiment 6, and 32.8 (SD = 11.4) for Experiment 7.

Materials

A total of 96 face images, each with a unique identity, were selected from the Chicago Face Database (CFD; Ma, Correll, & Wittenbrink, 2015). Faces were of Caucasian adults who were young to middle-aged. Half of the images were of women, and the other half were of men. All face images had a neutral expression, included only the individual's head and shoulders, and contained no facial accessories such as glasses or hats. All faces were presented in front view, in colour, on a white background.

Context stimuli consisted of 96 pictures of various scenes, presented in colour. Scenes were collected through an Internet search of websites that provided public access to their images. Scenes were selected to depict either high anxiety-provoking (e.g., a crowded party, a panel of interviewers) or low anxiety-provoking (e.g., an empty bedroom, a museum exhibit) scenarios.

Amongst the high anxiety-provoking scenes, half contained visible faces with discernible facial features embedded within the scene (henceforth referred to as embedded faces), and the other half contained no embedded faces. None of the low anxiety-provoking scenes contained any embedded faces. Various properties of the context scenes were assessed in a separate validation study in which naïve participants rated each scene in terms of anxiety/fear, avoidance, and visual complexity, as well as feelings of arousal and valence.

4.1.2 Scene Validation

Participants

One hundred naïve undergraduate students at the University of Waterloo were recruited for a 45-minute online study. Six participants were excluded from the final analyses due to missing responses for more than 40% of the items (two standard deviations above the mean; M_{missing} = 5.6%, SD_{missing} = 17.5%). One additional participant was excluded from the final analyses due to invariant responses to items (i.e., the same responses to 100% of items across all trials). Of the remaining 93 participants in the final analyses, 53% were women, 46% were men, and 1% were non-binary. Mean age was 21.0 (*SD* = 5.7).

Procedure

Participants were shown 120 scene images (validation included all 96 of the scene images from Experiment 6, plus 24 novel scene images used in Experiment 7) and made a series of judgments for each image. On each trial, participants were shown a single scene image (one at a time, in a randomized order) along with five questions below the image. The first two questions asked how anxious or fearful they felt in the scene (anxiety/fear), and how often they avoided the scene (avoidance) on 4-point Likert scales (items adapted from the LSAS; Liebowitz, 1987). The third question asked how visually complex the scene was, also on a 4-point Likert scale. The last two questions consisted of the arousal and valence subscales of the SAM (Bradley & Lang, 1994), rated on 9-point Likert scales.

Scene Validation Results

For each of the five scales, a 2 (Scene Type: low anxiety-provoking, high anxiety-provoking) × 2 (Scene Face Presence: without embedded faces, with embedded faces) repeated-measures ANOVA was conducted. Mean test scores are displayed in Table 4.

For the anxiety/fear scale, a significant main effect of Scene Type ($F(1, 92) = 220.37$, $MSE = 0.25$, $\eta_p^2 = .71$, $p < .001$) as well as a significant main effect of Scene Face Presence ($F(1, 92) = 33.12$, $MSE = 0.02$, $\eta_p^2 = .27$, $p < .001$) was found. In addition, the Scene Type by Scene Face Presence interaction was significant, $F(1, 92) = 11.78$, $MSE = 0.02$, $\eta_p^2 = .11$, $p < .001$. Post hoc paired *t*-tests revealed that the presence of faces significantly increased anxiety/fear ratings in high anxiety-provoking scenes ($t(92) = 6.49$, $SE = .02$, $p < .001$), but not in low anxiety-provoking scenes ($p = .05$).

For the avoidance scale, a significant main effect of Scene Type ($F(1, 92) = 174.79$, $MSE = 0.26$, $\eta_p^2 = .66$, $p < .001$) as well as a significant main effect of Scene Face Presence ($F(1, 92) = 8.25$, $MSE = 0.03$, $\eta_p^2 = .08$, $p = .005$) was found. In addition, the Scene Type by Scene Face Presence interaction was significant, $F(1, 92) = 10.02$, $MSE = 0.03$, $\eta_p^2 = .10$, $p = .002$. Post hoc paired *t*-tests revealed that the presence of faces significantly increased avoidance ratings in high anxiety-provoking scenes ($t(92) = 4.09$, $SE = .03$, $p < .001$), but not in low anxiety-provoking scenes ($p = .96$).

For the visual complexity scale, a significant main effect of Scene Type ($F(1, 92) = 112.18$, $MSE = 0.19$, $\eta_p^2 = .55$, $p < .001$) as well as a significant main effect of Scene Face Presence ($F(1, 92) = 45.82$, $MSE = 0.03$, $\eta_p^2 = .33$, $p < .001$) was found. In addition, the Scene

Type by Scene Face Presence interaction was significant, $F(1, 92) = 19.56$, $MSE = 0.02$, $\eta_p^2 = .18$, $p < .001$. Post hoc paired *t*-tests revealed that although the presence of faces significantly decreased visual complexity ratings in both low anxiety-provoking scenes ($t(92) = 7.04$, $SE = .03$, $p < .001$) and high anxiety-provoking scenes ($t(92) = 3.25$, $SE = .02$, $p = .002$), the magnitude of this decrease was greater for low anxiety-provoking scenes.

For the arousal scale, a significant main effect of Scene Type ($F(1, 92) = 191.75$, $MSE = 0.95$, $\eta_p^2 = .68$, $p < .001$) as well as a significant main effect of Scene Face Presence ($F(1, 92) = 14.01$, $MSE = 0.09$, $\eta_p^2 = .13$, $p < .001$) was found. In addition, the Scene Type by Scene Face Presence interaction was significant, $F(1, 92) = 27.42$, $MSE = 0.09$, $\eta^2_p = .23$, $p < .001$. Post hoc paired *t*-tests revealed that the presence of faces significantly increased arousal ratings in high anxiety-provoking scenes ($t(92) = 7.13$, $SE = .04$, $p < .001$), but not in low anxiety-provoking scenes ($p = .36$).

For the valence scale, a significant main effect of Scene Type ($F(1, 92) = 84.91$, $MSE = 0.74$, $\eta^2_p = .48$, $p < .001$) as well as a significant main effect of Scene Face Presence ($F(1, 92) = 26.98$, $MSE = 0.09$, $\eta^2_p = .23$, $p < .001$) was found. High anxiety-provoking scenes were rated as significantly more negative ($M = 5.44$) than low anxiety-provoking scenes ($M = 6.26$). As well, scenes with embedded faces were rated as significantly more negative ($M = 5.77$) than scenes without embedded faces ($M = 5.93$). No significant Scene Type by Scene Face Presence interaction was found ($p = .068$).

Table 4. *Experiments 6 and 7: Mean scene validation ratings for scenes without and with embedded faces (standard deviation in parentheses).*

Test Measure	Scene Without Embedded Faces	Scene With Embedded Faces
Fear		
Low Anxiety	0.30 (.31)	0.34 (.39)
High Anxiety	1.04 (.63)	1.16 (.66)
Avoidance		
Low Anxiety	0.53 (.46)	0.55 (.50)
High Anxiety	1.19 (.63)	1.28 (.68)
Visual Complexity		
Low Anxiety	0.66 (.45)	0.50 (.38)
High Anxiety	1.10 (.57)	1.04 (.58)
SAM Arousal		
Low Anxiety	2.69 (1.29)	2.67 (1.28)
High Anxiety	3.93 (1.54)	4.19 (1.61)
SAM Valence		
Low Anxiety	6.41 (1.10)	6.18 (1.18)
High Anxiety	5.52 (1.13)	5.42 (1.15)

Note: Valence was scored on a scale whereby higher scores corresponded with higher positivity.

Design

The 96 face images were randomly divided in half to form two study lists (List A and List B), each consisting of 48 faces (half female and half male). For half of the participants, faces from List A were presented as targets during encoding, whereas the other half saw faces from List B during encoding. Faces from the opposite list were used as lures during the face recognition phase. Lists were counterbalanced across participants. Similarly, the 96 context scenes were divided into two study lists (List 1 and List 2). Each list contained half high anxiety-provoking and half low anxiety-provoking contexts. Amongst the high anxiety-provoking contexts, half contained embedded faces and half did not. For half of the participants, scenes from List 1 were paired with target faces at encoding, whereas the other half saw scenes from List 2 at encoding. Scenes from the opposite list were used as novel context scenes at the face recognition phase. As with the face lists, context lists were counterbalanced across participants.

Procedure

Participants were recruited for a single 30-minute session and began by providing informed, written consent. Participants then completed an incidental encoding phase, in which they were presented with a total of 48 faces uniquely paired with 48 different context scenes. During each trial, participants were simultaneously presented a target face alongside a context scene. To ensure that participants were encoding both the face and the context incidentally, participants were asked to make two judgments for each face-context pair. First, participants rated the degree to which the person in the face image belonged in the context scene, from 1 (*very unlikely*) to 6 (*very likely*). Second, participants rated the degree to which they personally belonged in the context scene, from 1 (*very unlikely*) to 6 (*very likely*). Each judgment was presented as a Likert scale beneath the face-context pair (which remained visible). The face-context pair was presented for 2250 ms regardless of when the Likert scale ratings were made, in order to ensure equal encoding time across all face-context pairs. The Likert scales for the judgments were displayed for a maximum duration of 4000 ms. Following the response to the second judgment, a fixation cross was displayed in the center of the screen for 500 ms before proceeding to the next trial. All 48 face-context pairs were presented sequentially in a randomized order. Of the context scenes, half depicted high anxiety-provoking scenarios and half depicted low anxiety-provoking scenarios.

Following the encoding phase, participants completed the recognition phase for the target faces. Ninety-six face-context pairs were sequentially presented in a randomized order. For each trial, participants made a recognition decision indicating whether the face displayed was old or new using a Remember/Know/New (R/K/N; Tulving, 1985) paradigm. To indicate that a face was old, participants pressed a key labelled either “R” or “K”, representing “remember” or “know”. Alternatively, participants pressed a key labelled “N” to indicate the face was “new” and not from the study set. Once a response was made, the next face-context pair was presented. Forty-eight of the face-context pairs contained old faces, 24 of which were presented alongside the exact same context image as during the encoding phase (i.e., context was reinstated). The other 24 old faces were paired with novel context scenes of the equivalent context type (i.e., context was not reinstated). For example, a target face encoded with a low anxiety-provoking scene would be paired with a new, low anxiety-provoking scene at retrieval. The remaining 48 face-context pairs contained new faces (i.e., lure face-context pairs), and were intermixed amongst the 48 face-context pairs containing old faces. Twenty-four of these lure face-context pairs contained old context scenes, while the other 24 contained new context scenes.

After completing the face recognition phase, participants completed a scene recognition phase. Ninety-six context scenes were presented alone in the center of the screen in a randomized order. For each trial, participants made a recognition decision using the same R/K/N paradigm as the face recognition phase. Of the 96 context scenes shown, 48 were studied during the encoding phase (old), and the other 48 were never presented previously (new). For both the old context scenes and the new context scenes, 24 were low anxiety-provoking and 24 were high anxiety-provoking. Finally, all participants completed an in-person, computerized administration of the SPIN (Connor et al., 2000) as a measure of trait social anxiety (SA). All study procedures were approved by the Office of Research Ethics at the University of Waterloo (Protocol #30730).

4.1.2 Results

Face Recognition Accuracy

Hit rate was tabulated for each trial type and for each scene type, in each participant, by summing the number of correct R and K responses and dividing by 48. False alarm rate was tabulated for each participant by summing the number of incorrect R and K responses and dividing by 48 (see Table 5 for means). Accuracy for each participant in each condition was calculated by subtracting an individual's false alarm rate from their hit rate. Mean accuracy rates are displayed in Figure 6.

Accuracy data were analyzed in a 2 (Trial Type: paired with new scene, paired with old scene) × 3 (Scene Type: low anxiety-provoking without faces, high anxiety-provoking without faces, high anxiety-provoking with faces) repeated-measures ANOVA. There was a significant main effect of Trial Type, $F(1, 48) = 23.82$, $MSE = 0.06$, $\eta_p^2 = .33$, $p < .001$, such that faces paired with reinstated contexts had significantly higher accuracy ($M = .52$) than faces paired with new contexts ($M = .39$). The main effect of Scene Type was not significant ($p = .52$), but the

Trial Type × Scene Type interaction was, $F(2, 96) = 7.54$, $MSE = 0.04$, $\eta_p^2 = .14$, $p = .001$. Specifically, memory was enhanced for faces paired with old ($M = .58$) compared to new ($M = .37$) low anxiety-provoking scenes without embedded faces, $t(48) = 6.29$, $SE = 0.03$, $p < .001$. In a similar pattern, memory was also enhanced for faces paired with old ($M = .53$) compared to new ($M = .35$) high anxiety-provoking scenes without embedded faces, $t(48) = 3.93$, $SE = 0.05$, $p < .001$. Importantly, memory was not significantly different between faces paired with old ($M = .45$) or new ($M = .44$) high anxiety-provoking scenes with embedded faces, $t(48) = 0.15$, $SE = 0.05$, $p = .89$.

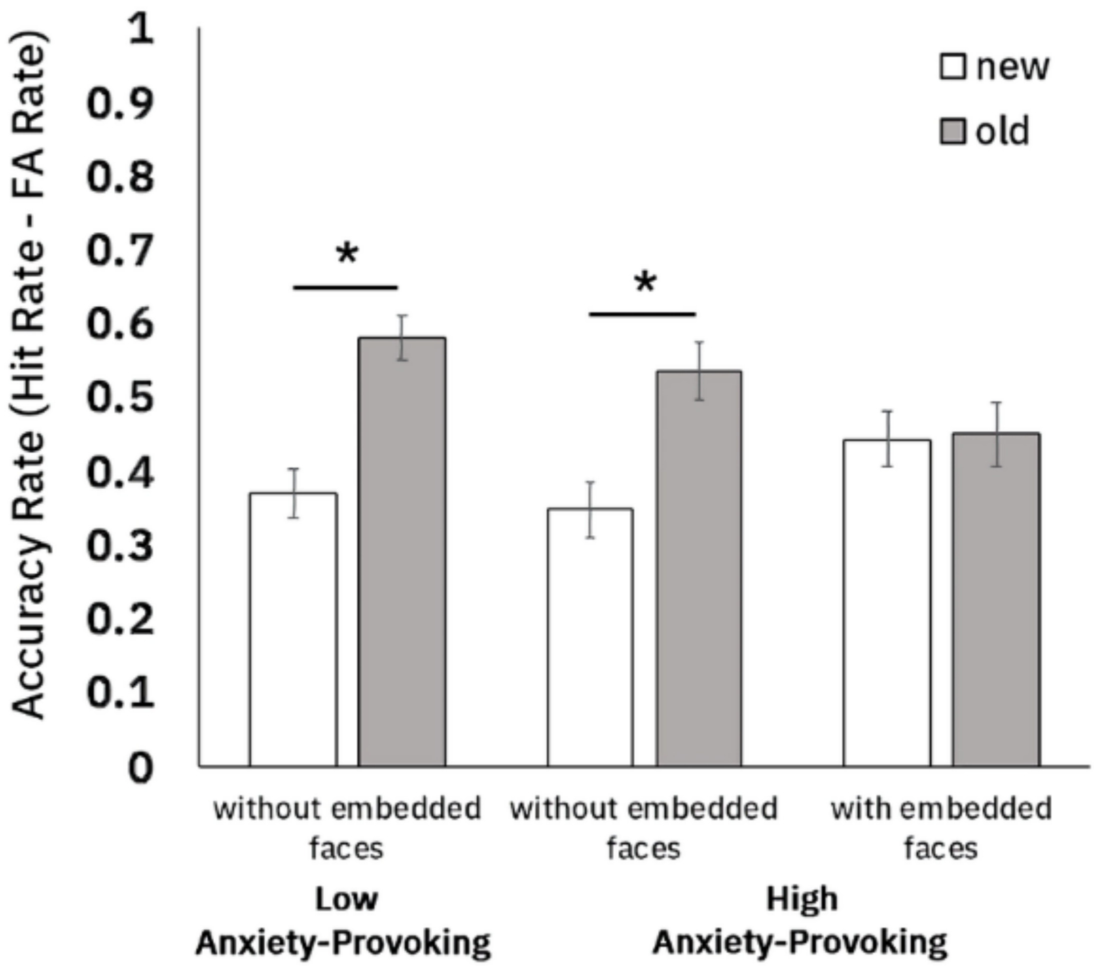

Figure 6. Accuracy rate (hit rate – false alarm rate) for faces paired with low anxiety-provoking and high anxiety-provoking (without and with embedded faces) context scenes.

Scene Recognition Accuracy

Hit rate was tabulated for each scene type, in each participant, by summing the number of correct R and K responses and dividing by 48. False alarm rate was tabulated for each participant by summing the number of incorrect R and K responses and dividing by 48 (see Table 6 for means). Accuracy for each participant in each scene type was calculated by subtracting an individual's false alarm rate from their hit rate.

A one-way repeated-measures ANOVA was conducted to compare the effect of Scene Type (low anxiety-provoking without faces, high anxiety-provoking without faces, high anxiety-provoking with faces) on memory for the scenes. Due to Mauchly's test indicating that the assumption of sphericity had been violated, $\chi^2(2) = 6.07$, p = .048, a Greenhouse-Geisser correction was applied when determining significance. I found a significant main effect of Scene Type, $F(1.78, 85.62) = 23.71$, $MSE = 0.02$, $\eta_p^2 = .33$, $p < .001$. Post hoc comparisons found that memory for high anxiety-provoking scenes without embedded faces was significantly poorer ($M = .42$) than both low anxiety-provoking scenes ($M = .59$; $t(48) = 7.85$, $SE = 0.02$, $p < .001$) and high anxiety-provoking scenes with embedded faces ($M = .54$; $t(48) = 4.93$, $SE = 0.03$, $p < .001$), which were not significantly different from each other, $t(48) = 1.60$, $SE = 0.03$, $p = .12$.

Correlations with Scene Recognition and Face Recognition

Three correlations were conducted between memory accuracy for scenes and memory accuracy for faces: one for each of the three scene types (low anxiety-provoking without faces, high anxiety-provoking without faces, high anxiety-provoking with faces). All three correlations were non-significant (*p*s > .51), suggesting that participants' ability to accurately remember a face seen in a given context type was not related to their ability to accurately remember its paired context scene.

Correlations with Trait Social Anxiety.

Face Recognition. Three correlations were conducted between SPIN score and memory accuracy for target faces: one for each of the three scene types (low anxiety-provoking without faces, high anxiety-provoking without faces, high anxiety-provoking with faces). A significant positive correlation was found between SPIN score and memory accuracy for faces that were paired with high anxiety-provoking scenes without embedded faces ($r = .32$, $p = .02$). The correlations between SPIN score and memory accuracy for faces paired with low anxiety-provoking scenes ($p = .20$) and high anxiety-provoking scenes with embedded faces ($p = .47$) were non-significant.

Scene Recognition. Three correlations were conducted between SPIN score and memory accuracy for scenes: one for each of the three scene types (low anxiety-provoking without faces, high anxiety-provoking without faces, high anxiety-provoking with faces). A significant negative correlation was found between SPIN score and memory accuracy for high anxiety-provoking scenes without embedded faces ($r = -.30$, $p = .03$). The correlations with memory accuracy for faces paired with low anxiety-provoking scenes ($p = .23$) and high anxiety-provoking scenes with embedded faces ($p = .32$) were non-significant.

4.1.3 Discussion

In the current experiment, I investigated how the anxiety-provoking nature of a context might influence that context's ability to provide a memory benefit via the CR effect. Specifically, I predicted that high anxiety-provoking contexts would reduce the CR effect, whereas low anxiety-provoking scenes would still provide the classic benefit to memory. This was hypothesized given that some past reports have shown that contexts which evoke common components of anxiety (negative valence and high arousal) reduce the typical CR benefit

(Brown, 2003; Rainis, 2001). The current results partially supported this hypothesis. I did find that the CR effect was eliminated for high anxiety-provoking contexts – however, this was only true for the high anxiety-provoking contexts with embedded faces. In contrast, the high anxiety-provoking contexts without embedded faces showed the classic CR effect, much like the low anxiety-provoking contexts. It would therefore seem that a critical feature of the contexts driving the reduction in CR effect is the presence of embedded faces in the context.

Importantly, the current evidence also rules against the possibility that the reduction in CR effect could be explained by a trade-off between face memory and scene memory. Past work suggests that high anxiety-provoking contexts could capture attention (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & Van Ijzendoorn, 2007; Cisler & Koster, 2010; Mogg et al., 2000) and therefore be processed locally in detail, at the detriment of globally binding the target face to the context. A trade-off account would therefore expect that since high anxiety-provoking contexts with embedded faces showed no CR effect, we would also see a corresponding benefit to scene memory. This was not the case in the current experiment – memory for high anxiety-provoking scenes with embedded faces were not significantly different from low anxiety-provoking scenes. As such, the findings suggest that the observed reduction in CR effect is not due to selective processing of the high anxiety-provoking scene.

4.2 Experiment 7

Results of Experiment 6 suggest that highly anxiety-provoking scenes fail to confer the expected memory benefit when they are reinstated. Specifically, a reduction in CR benefit was found for target faces paired with high anxiety-provoking scenes with embedded faces. A limitation of this experiment, however, is that the other context types (low anxiety-provoking scenes, and high anxiety-provoking scenes without faces) could not allow us to determine whether the reduced CR effect was due to the presence of faces, or the anxiety-provoking nature of the scene itself. To identify which of these features led to the reduced CR effect, I added an additional context type in Experiment 7 – low anxiety-provoking context scenes containing embedded faces. If the presence of faces drove the reduction in CR benefit, we would expect to see both the low anxiety-provoking and high anxiety-provoking scenes with faces offer diminished CR benefit relative to scenes without faces. Alternatively, the reduction in CR benefit may have been driven by the combination of embedded faces in the context scene, as well as the highly anxiety-provoking quality of the scene. If this were the case, we would predict a reduced CR benefit from high anxiety-provoking scenes with faces, in comparison to high anxiety-provoking scenes without faces or low anxiety-provoking scenes, regardless of whether embedded faces are present.

4.2.1 Method

Participants

Fifty undergraduate students who had not participated in Experiment 6 completed the current experiment for partial course credit. All recruitment procedures were identical to that of the previous experiment. Of the current experiment's participants, 82% were female and 18% were male, and the mean age was 19.8 (*SD* = 2.0).

Materials

All materials were identical to that of the previous experiment, except for the addition of a new context scene type: low anxiety-provoking scenes with embedded faces.

Context stimuli were identical to that of the previous experiment, except that 24 of the 48 low anxiety-provoking scenes from the previous experiment were replaced with 24 new, low anxiety-provoking scenes with embedded faces. These new scenes were selected using the same procedure as Experiment 6. As well, the replaced scenes from Experiment 6 were randomly selected from the low anxiety-provoking scenes without embedded faces. In sum, the current experiment contained 24 low anxiety-provoking scenes without embedded faces, 24 low anxiety-provoking scenes with embedded faces, 24 high anxiety-provoking scenes without embedded faces, and 24 high anxiety-provoking scenes with embedded faces. For any given participant, 12 of the 24 scenes of each type were seen as targets in the encoding phase, whereas the other 12 were seen as lures in the recognition phases.

Design and Procedure

The design was identical to that of the previous experiment, except for the addition of low anxiety-provoking scenes with embedded faces. In other words, the anxiety-provoking nature of the scene and the presence of embedded faces were fully disentangled in the current experiment. Specifically, 12 target faces were paired with low anxiety-provoking scenes without embedded faces, 12 were paired with low anxiety-provoking scenes with embedded faces, 12 were paired alongside high anxiety-provoking scenes without embedded faces, and 12 were paired with high anxiety scenes with embedded faces. The procedure was identical to the previous experiment.

4.2.2 Results

Face Recognition Accuracy

Accuracy was calculated using the same procedure as Experiment 6 and is summarized in Table 5 and Figure 7. Accuracy data were analyzed in a 2 (Trial Type: paired with new scene, paired with old scene) × 2 (Scene Anxiety Type: low anxiety-provoking, high anxiety-provoking) × 2 (Scene Face Presence: without faces embedded in scene, with faces embedded in scene) repeated-measures ANOVA.

Similar to Experiment 6, there was a significant main effect of Trial Type, $F(1, 49) = 7.78$, $MSE = 0.07$, $\eta_p^2 = .14$, $p = .008$, such that faces paired with reinstated contexts had significantly higher accuracy ($M = .47$) than faces paired with new contexts ($M = .39$). The main effect of Scene Anxiety Type was close to significance ($p = .052$), such that faces paired with low anxiety-provoking contexts showed nominally greater accuracy ($M = .46$) than face paired with high anxiety-provoking contexts ($M = .41$). The main effect of Scene Face Presence was not significant ($p = .18$).

Also replicating Experiment 6, I found a significant Trial Type × Scene Anxiety Type interaction, $F(1, 49) = 12.15$, $MSE = 0.06$, $\eta_p^2 = .20$, $p = .001$. Specifically, memory was enhanced for faces paired with old ($M = .53$) compared to new ($M = .38$) low anxiety-provoking scenes, $t(49) = 4.40$, $SE = 0.04$, $p < .001$. In contrast, memory was not significantly different between faces paired with old ($M = .40$) or new ($M = .41$) high anxiety-provoking scenes, $t(49) = 0.19$, $SE = 0.04$, $p = .85$.

Additionally, a significant Trial Type × Scene Face Presence interaction was found, $F(1, 49) = 5.67$, $MSE = 0.04$, $\eta_p^2 = .10$, $p = .02$. Specifically, memory was enhanced for faces paired with old ($M = .47$) compared to new ($M = .35$) scenes without embedded faces, $t(49) = 3.46$, $SE = 0.04$, $p = .001$. In contrast, memory was not significantly different between faces paired with old ($M = .46$) or new ($M = .43$) scenes with embedded faces, $t(49) = 0.90$, $SE = 0.03$, $p = .37$.

A significant three-way (Trial Type × Scene Anxiety Type × Scene Face Presence) interaction was also revealed, $F(1, 49) = 6.35$, $MSE = 0.05$, $\eta_p^2 = .12$, $p = .02$. To examine this three-way interaction further, an additional 2 (Trial Type: paired with new scene, paired with old scene) × 2 (Scene Anxiety Type: low anxiety-provoking, high anxiety-provoking) repeated-measures

ANOVA was conducted at each level of Scene Face Presence (without faces embedded in scene, with faces embedded in scene).

For scenes without embedded faces, a significant main effect of Trial Type was found, $F(1, 49) = 11.97$, $MSE = 0.06$, $\eta_p^2 = .20$, $p = .001$, such that faces paired with old scenes (M = .47) were more accurately remembered than faces paired with new scenes (M = .35). The main effect of Scene Anxiety Type was not significant ($p = .06$). Similarly, the Trial Type × Scene Anxiety Type interaction was not significant ($p = .40$), indicating that I was unable to find evidence that Scene Anxiety Type had an effect on CR for scenes without embedded faces.

For scenes with embedded faces, no significant main effects were found of either Trial Type (p = .37) or Scene Anxiety Type ($p = .38$). However, the Trial Type by Scene Anxiety Type interaction was significant, $F(1, 49) = 6.35$, $MSE = 0.05$, $\eta_p^2 = .12$, $p < .001$, indicating that Scene Anxiety Type influenced CR for scenes with embedded faces. Specifically, memory was enhanced for faces paired with old ($M = .55$) compared to new ($M = .38$) low anxiety-provoking scenes with embedded faces, $t(49) = 3.26$, $SE = 0.05$, $p = .002$. In contrast, memory was significantly impaired for faces paired with old ($M = .38$) compared to new ($M = .49$) high anxiety-provoking scenes with embedded faces, $t(49) = 2.57$, $SE = 0.04$, $p = .01$.

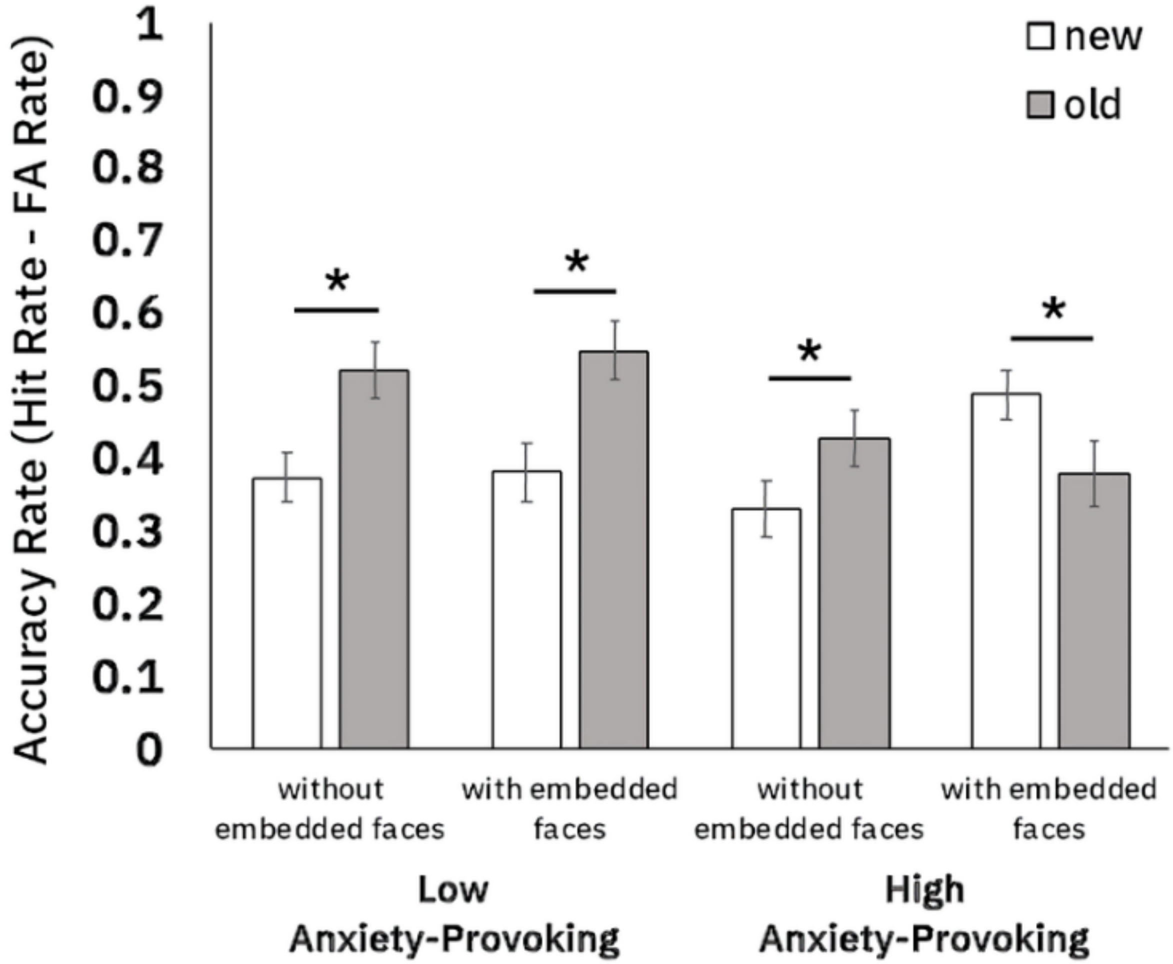

Figure 7. Accuracy rate (hit rate – false alarm rate) for faces paired with either low, or high, anxiety-provoking scenes, subdivided into context with and without embedded faces.

Scene Recognition Accuracy.

Scene memory accuracy was calculated for each participant using the same method as Experiment 6 (see Table 6 for mean accuracy rates). A 2 (Scene Anxiety Type: low anxiety-provoking, high anxiety-provoking) × 2 (Scene Face Presence: without faces embedded in scene, with faces embedded in scene) repeated-measures ANOVA was conducted to compare memory performance.

No significant main effects were found for Scene Anxiety Type ($p = .07$) or Scene Face Presence ($p = .66$). However, a significant Scene Anxiety Type × Scene Face Presence interaction was found, $F(1, 49) = 10.62$, $MSE = 0.02$, $\eta_p^2 = .18$, $p = .002$. Post hoc paired t-tests revealed that memory accuracy was higher for low anxiety-provoking scenes with embedded faces ($M = .49$) relative to those without embedded faces ($M = .43$; $t(49) = 2.51$, $SE = 0.03$, $p = .02$). Conversely, memory accuracy was higher for high anxiety-provoking scenes without embedded faces ($M = .46$) relative to those with embedded faces ($M = .38$; $t(49) = 2.64$, $SE = 0.03$, $p = .01$).

Correlations with Scene Recognition and Face Recognition

Four correlations were conducted between memory accuracy for scenes and memory accuracy for faces: one for each of the two scene anxiety types (low anxiety-provoking, high anxiety-provoking), at each of the two levels of scene face presence (with faces embedded in scene, without faces embedded in scene). All four correlations were non-significant (p's > .08), replicating the finding from Experiment 6 that participants' ability to accurately remember a face seen in a given context type was not related to their ability to accurately remember its paired context scene.

Correlations with Trait Social Anxiety

Face Recognition. Four correlations were conducted between SPIN score and memory accuracy for target faces: one for each of the two scene anxiety types (low anxiety-provoking, high anxiety-provoking), at each of the two levels of scene face presence (with faces embedded in scene, without faces embedded in scene). All four correlations were non-significant (p's > .14).

Scene Recognition. Four correlations were conducted between SPIN score and memory accuracy for scenes: one for each of the two scene anxiety types (low anxiety-provoking, high anxiety-provoking), at each of the two levels of scene face presence (with faces embedded in scene, without faces embedded in scene). All four correlations were non-significant (p's > .43).

4.2.3 Discussion

The current experiment aimed to specify the context features that might be driving the observed reductions in CR. By including an additional context type (low anxiety-provoking with embedded faces), the current experiment fully disentangled the anxiety-provoking nature of the scenes from the presence of visible faces in the scenes. My results showed that the combination of both scene anxiety type and scene face presence was critical in reducing the CR effect.

First, I replicated findings from Experiment 6 in that both low anxiety-provoking scenes without embedded faces and high anxiety-provoking scenes without embedded faces showed the expected CR effect: memory for faces was enhanced when their paired scenes were reinstated. Critically, I also replicated the key finding from Experiment 6, where high anxiety-provoking scenes with embedded faces failed to confer the classic CR benefit – further, this effect was even stronger in Experiment 7, such that reinstating these scenes significantly impaired memory for faces. The novel contribution of Experiment 7 was the finding that this elimination or reversal of the CR effect was not observed with the new scene type of low anxiety-provoking scenes with embedded faces. Much like the low anxiety-provoking scenes without embedded faces or the high anxiety-provoking scenes with embedded faces, these new scenes showed the typical CR effect (see Figure 7). This finding argues that the presence of embedded faces in a scene is not sufficient to reduce the CR effect – rather, it appears that the combination of embedded faces and a high anxiety-provoking scenario is needed to influence CR.

Importantly, I also replicated the lack of any observed trade-off between face memory and scene memory found in Experiment 6. A selective attention hypothesis calling for preferential encoding of highly anxiety-provoking scenes, rather than their paired target faces, would necessarily predict a trade-off in recognition performance. Specifically, this hypothesis would be supported if scene recognition and face recognition were negatively correlated, such that scene memory benefited at the expense of face memory. However, no evidence for a trade-off was observed in Experiment 7: correlations were non-significant between face and scene memory across all scene types. Therefore, as in Experiment 6, it is unlikely that the pattern of results observed is due to the selective processing of either the target faces or their accompanying context scenes.

4.3 General Discussion

In the current series of experiments, I examined whether semantic features of a context affect the memory benefit conferred by reinstating the encoding context at retrieval. In Experiment 6, the expected CR benefit was observed when the context consisted of low anxiety-provoking scenes, and high anxiety-provoking scenes without embedded faces. In contrast, the CR benefit was significantly reduced when the contexts were high anxiety-provoking scenes containing embedded faces. In Experiment 7, I included an additional context type, consisting of low anxiety-provoking scenes with embedded faces. Once again, the CR effect was shown to be significantly reduced only when the context scenes were high anxiety-provoking with embedded faces: reinstating the same, versus new, context failed to enhance memory for targets. Results suggest that the benefit of reinstating a context, on target memory, depends critically on semantic characteristics of the reinstated context.

Across both experiments, the reduction in CR only occurred when the context was highly anxiety-provoking and contained embedded faces. In other words, it was only when the context involved both an anxiety-provoking scenario and had visible faces that the CR effect was eliminated; neither component alone eliminated the CR effect. Why might the combined effects of scene anxiety type and scene face presence drive this reduction in CR? Evidence from the scene validation study suggests that the reduction in CR may stem from emotional reactions to the semantic features of a context (e.g., whether or not a scene is anxiety-provoking or not). In particular, my validation study revealed that the context images involving both high anxiety-provoking scenarios and embedded faces induced greater feelings of anxiety/fear, avoidance, and arousal compared to any other scene type.

One potential explanation for why these negative feelings could reduce CR effects is that the emotional reactions to high anxiety-provoking scenes with embedded faces could lead to preferential processing of the context scene, to the detriment of encoding the target face. For instance, highly anxiety-provoking contexts with embedded faces could be capturing attention (Bar-Haim et al., 2007; Cisler & Koster, 2010; Mogg et al., 2000) due to the threat-related feelings evoked by these specific scenes. As such, one might argue that participants were selectively encoding the highly anxiety-provoking scenes, instead of processing the target faces within the context of the anxiety-provoking scenes. This hypothesis would necessarily predict a trade-off between context memory and target memory if attention is being directed towards the scenes and away from the faces – that is, as memory for emotionally arousing or anxiety-provoking scenes improves, memory for faces paired with them ought to decrease. However, this pattern was not observed in my data: memory for high anxiety-provoking scenes with embedded faces was not correlated with memory for the target faces paired with those scenes in either of the current experiments. It would therefore seem that a selective attention account is not sufficient to fully explain my findings.

An alternative hypothesis is that the negative feelings evoked by high anxiety-provoking scenes with embedded faces could reduce the CR effect by interfering with the ability to bind target faces to their paired contexts. In other words, rather than impairing memory for one item or the other in a face-context pair, the emotional reactions could have hindered the connections between faces and their contexts in memory. Specifically, the arousal-biased competition model claims that emotional arousal should impair associations between an item and its context when explicit item-context binding is not a high-priority task (Mather & Sutherland, 2011). Given that the encoding of items within their contexts was incidental in the current experiments, arousal-biased competition would predict that the emotionally arousing scenes (i.e., high anxiety-provoking with embedded faces) would lead to weaker associations between the target faces and their paired context scenes. Critically, the memory benefits typically conferred by CR are thought to rely on integrated memory traces – that is, that CR boosts memory when the context is encoded together with the target items (Smith & Vela, 2001), and that CR fails to enhance memory if these strong associations are not formed (Murnane, Phelps, & Malmberg, 1999). The high anxiety-provoking scenes with embedded faces could therefore be decreasing the strength with which contextual information is associated with the target faces, causing the observed lack of CR effects with this scene type.

In addition to elaborating upon the mechanisms by which emotion interacts with the CR effect, my results may also have implications for the literature on eyewitness memory. Replicating past research, my findings demonstrate that negative or emotionally arousing scenes reduce the boost to memory typically offered by reinstating a context (Brown, 2003; Rainis, 2001). This failure to enhance memory may be particularly troubling for suspect identification, as many crime scenes are inherently negative in valence or highly anxiety-provoking. However, there is research demonstrating benefits to reinstating contexts in crime-related situations (Hammond, Wagstaff & Cole, 2006; Krafka & Penrod, 1985; Wong & Read, 2011). Given the importance and consequences of eyewitness identification, future work should continue to investigate the boundary conditions of reinstating anxiety-provoking contexts, so as to improve the accuracy of eyewitness identification.

Another factor that may have influenced CR effects in the current work is an individual's trait level of social anxiety. Evidence suggests that one's level of trait social anxiety is related to altered processing of socially threatening information across many domains of cognition, such as attention (Bar-Haim et al., 2007; Cisler & Koster, 2010; Mogg et al., 2000; Shechner et al., 2012) and memory (Herrera, Montorio, Cabrera, & Botella, 2017; Mitte, 2008; Yeung & Fernandes, 2019a, 2019b). In the current experiments, I measured trait social anxiety using the SPIN (Connor et al., 2000) to investigate whether level of trait social anxiety would further influence the CR benefit offered by the context scenes. Specifically, it was possible that participants high in trait social anxiety would experience more intense emotional reactions to the anxiety-provoking context scenes than those low in trait social anxiety. In turn, these stronger feelings of emotional arousal could possibly reduce the CR effect even further if one is higher as opposed to lower in trait social anxiety. However, we did not find consistent, systematic correlations between trait social anxiety and memory accuracy (in the form of face memory as well as scene memory) across both experiments.

Because memory performance was not significantly related to trait social anxiety, regardless of scene anxiety type or scene face presence, my findings suggest that a general propensity to experience social anxiety does not influence the reduction in CR effects. Nevertheless, it remains possible that I may have failed to observe associations between memory performance and trait social anxiety due to the specific stimuli used in the current experiment. To make sure that participants would consistently judge the highly anxiety-provoking scenes as anxiety-provoking, I selected scenes of situations that are very commonly feared by the general population (e.g., preparing to give a presentation or being interviewed). Although I may have ensured that participants would experience emotional reactions to the scenes, this may have blunted my ability to detect differences across levels of trait SA. Given that the highly anxiety-provoking scenes were so unambiguously threatening, any participant – regardless of trait SA – may have felt equally strong emotional arousal in response to the highly anxiety-provoking scenes. Indeed, past findings contend that group differences between those low or high in trait SA are sometimes only observed when ambiguously threatening stimuli rather than universally threatening stimuli (Amir, Beard, & Bower, 2005; Constans, Penn, Ihen, & Hope, 1999). Future work should consider varying the anxiety-provoking nature of scenes with more granularity in order to further probe the influence of trait SA on the CR effect.

4.3.1 Conclusions

Across two experiments, I observed a reduction in the CR effect when target faces were studied within the context of highly anxiety-provoking scenes. In other words, although reinstating the context from encoding typically enhances memory (Smith & Vela, 2001), high anxiety-provoking contexts with embedded faces failed to improve recognition when reinstated in the current experiments. As evidenced by the lack of a trade-off between face and scene recognition accuracy, the key findings support an arousal-biased competition account (e.g., emotional arousal impairing the binding of low-priority inter-item associations) rather than a selective attention account (e.g., preferential encoding of the emotionally arousing context) of how emotion may influence the CR effect. My work adds to the growing list of contextual features (Baddeley, 1982; Dalton, 1993; Skinner & Fernandes, 2010; Smith & Vela, 2001) which may reduce the CR effect. Further, I extend past research by showing that the combined effects of anxiety-provoking situations and embedded faces appear to be necessary components to reducing the CR effect when considering socially relevant scenes as contexts.

Table 5

Hit, false alarm (FA), and accuracy rates for face recognition in Experiments 6 and 7 as a function of recognition test trial type, scene anxiety type, and scene face presence (standard deviations in parentheses).

Experiment 6

	Low Anxiety-Provoking			**High Anxiety-Provoking**					
	Without Embedded Faces			**Without Embedded Faces**			**With Embedded Faces**		
Trial Type	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate
New	.57 (.20)	.21 (.17)	.37 (.23)	.54 (.27)	.19 (.18)	.35 (.26)	.63 (.25)	.18 (.22)	.44 (.26)
Old	.78 (.17)	.20 (.19)	.58 (.21)	.75 (.23)	.22 (.22)	.53 (.27)	.67 (.26)	.22 (.20)	.45 (.31)

Experiment 7

	Low Anxiety-Provoking						**High Anxiety-Provoking**					
	Without Embedded Faces			**With Embedded Faces**			**Without Embedded Faces**			**With Embedded Faces**		
Trial Type	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate
New	.57 (.25)	.20 (.20)	.37 (.23)	.56 (.24)	.18 (.20)	.38 (.28)	.57 (.25)	.24 (.22)	.33 (.28)	.69 (.19)	.21 (.19)	.49 (.24)
Old	.85 (.16)	.33 (.23)	.52 (.27)	.81 (.18)	.26 (.26)	.55 (.29)	.76 (.19)	.33 (.25)	.43 (.27)	.68 (.27)	.30 (.25)	.38 (.32)

Table 6

Hit, false alarm (FA), and accuracy rates for scene recognition in Experiments 6 and 7 as a function of trial type, scene anxiety type, and scene face presence (standard deviations in parentheses).

Experiment 6

Low Anxiety-Provoking			High Anxiety-Provoking					
Without Embedded Faces			Without Embedded Faces			With Embedded Faces		
Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate
.67 (.28)	.08 (.09)	.59 (.28)	.56 (.31)	.15 (.15)	.42 (.28)	.67 (.26)	.13 (.14)	54 (.23)

Experiment 7

Low Anxiety-Provoking						High Anxiety-Provoking					
Without Embedded Faces			With Embedded Faces			Without Embedded Faces			With Embedded Faces		
Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate	Hit Rate	FA Rate	Accuracy Rate
.89 (.10)	.46 (.22)	.43 (.22)	.91 (.11)	.42 (.24)	.49 (.22)	.86 (.14)	.40 (.24)	.46 (.23)	.89 (.10)	.51 (.22)	.38 (.22)

Chapter 5: General Discussion of PhD Experiments

Throughout a series of experiments, I investigated the influence of a variety of factors on context effects on recognition memory. Through Experiments 1 to 5, I identified two distinct patterns in memory performance. Specifically, the magnitude of the CR effect on words was unaffected by familiarity and was consistently large regardless of number of pre-exposures. A similar pattern was observed for target objects. Though a single pre-exposure did decrease CR magnitude in comparison to 0 pre-exposures, further increasing the degree of familiarity (number of pre-exposures) did not systematically reduce the magnitude of the CR effect. Such patterns of results are in line with the global matching models (TODAM2, Murdock, 1997; ICE model, Murnane & Phelps, 1999) that suggest that target-context pairs are encoded and retrieved as ensembles. As such, both parts of the pair will always contribute mathematically to recognition memory decisions.

A very different pattern was observed when faces, as opposed to words or objects, were used as the target stimulus. In line with the outshining hypothesis, as the number of pre-exposures to the face increased, the benefit of context reinstatement decreased. The outshining hypothesis suggests that the target and context act as cues for one another, each possessing its own independent memory signal. As the memory signal for one side of the pair becomes stronger, it eventually no longer requires the presence of the other side in order to inform memory decisions. Instead, the more robustly encoded side of the pairing outshines the other side. In my experiments, this is very much what was observed with faces. Faces that were very familiar (5 or 10 pre-exposures) developed a stronger memory signal, relative to the weaker signal conveyed by the context. Therefore, the signal for the face outshone that of the context, making the context's presence unnecessary for face recognition.

A key strength of these experiments is the consistency of many of the effects. As discussed in the introduction of this thesis, there has been a large amount of variability in the ways the CR effect has been investigated in the past. This includes different test materials (e.g. faces, objects, words), context types (e.g. pictures scenes, word pairings, screen background colour, experiment rooms, researchers) and even test type (recognition and recall). In Experiments 1 to 5, I investigated how familiarity with items, objects and faces can influence the CR effect. Importantly, I maintained other manipulations in across these experiments. That is, each experiment used the same set of scenes as context, and used the same type of retrieval test, recognition.

Additionally, familiarity with each of the stimulus types was also manipulated in the same way. In each experiment, I used number of pre-exposures to make some stimuli more familiar (5 or 10 pre-exposures), or less familiar (0 pre-exposures), to participants. This is important because in past work in the field, the familiarity manipulation from experiment to experiment has varied and often involved a contrast between qualitative categories (e.g. famous and non-famous people; real or non-real words). The issue with comparing such categories is that it becomes difficult to determine the *degree* to which each participant is actually familiar with each famous person. For example, a huge fan of film may be more familiar with a particular actor (i.e. face) than a participant who does not follow the film industry very closely. Therefore, such a dichotomy is problematic because it assumes a certain level of familiarity with famous faces. In these experiments, I selected a pre-exposure manipulation to avoid issues with prior knowledge and familiarity. Instead, I was able to *quantify* familiarity levels and then contrast these groups to one another. Moreover, as has already been touched upon, the experiments in this thesis all used the same familiarity manipulation. In this way, I was able to keep the familiarity manipulation, the context type and the memory test format consistent across studies, leaving the target stimulus type as the only factor that differed across studies. In this way, I could truly determine that the source of any differences in recognition memory performance was indeed due to the target stimulus type. Indeed, I did find that CR magnitude was uniquely affected by face, but not word or object, familiarity.

In Experiment 5 from Chapter 3, I extended the results of the previous experiments by, first, replicating the two patterns in a within-groups (Experiment 5) design. This replication provides support and converging evidence that it was truly the target stimulus type driving the two patterns of results and not due to different participant groups across experiments. Importantly, Experiment 5 also extended the previous results by examining whether overt attention could offer an explanation for the two patterns. I used eye tracking as an indirect measure of attention and measured both the proportion of fixations and dwell time on either the target (word or face) or the accompanying context at encoding. In this way, I could determine whether familiarity with either faces or words would influence the way in which participants attended to a target-context pair. Critically, I observed two distinct patterns in eye gaze behaviours between words and faces. For words, across all pre-exposure conditions, there were higher proportions of both fixation count and dwell time on the context, relative to the target word. In stark contrast, when faces were the study stimulus, there were higher proportions of fixation and dwell time on the face. These distinct patterns of eye gaze behaviours suggest that participants treat faces and words very differently during encoding. That is, faces but not words, relative to a context, seem to demand a higher proportion of attention..

The eye gaze proportion patterns are perhaps unsurprising given the number of details and features that are present in a face in comparison to a word. Indeed, when processing a face, there are the eyes, the nose, and the mouth amongst a number of other features that define a visage. A word, on the other hand, is quickly processed and may require less time to fully grasp. However, this short attendance on the word seems to result in memorial consequences. Taking the eye gaze and memory performance patterns in conjunction with one another suggests that the preference to attend to the context rather than target word at encoding ultimately results in dependence for the word, on the presence of the context to inform its memory. Critically, this dependence also seems to be the case regardless of familiarity or number of pre-exposures to a particular word. In contrast, faces, which receive a higher proportion of eye gaze, receive a reduced context benefit as familiarity increases. This pattern suggests that, in general, if a target item, such as a face, demands a higher proportion of attentional resources, relative to an associated context scene that is simultaneously presented, then relative familiarity with the target will influence the magnitude of benefit conferred by reinstating the context again at retrieval. However, if the target (e.g. a word in Experiments 1 and 5) does not demand a higher proportion of attentional resources, then the context may shield the target from any benefits of familiarity, which act independently of the context. That is, familiarity will have a reduced impact on word recognition because the context denies familiarity from playing such a role. Instead, the presence of the context becomes the dominating factor and integral cue in aiding memory performance.

In Experiments 6 and 7, I switched the focus to manipulating features of the context, instead of the target stimulus. In these studies, I found that highly social anxiety provoking context scenes that contain embedded faces provide reduced context benefit to face recognition. Across the series of studies within this thesis, I have provided evidence that a number of factors and features related to both the target item and the context influence the benefit of context effects. Not only does the type of target stimulus influence context effects, but also familiarity with the item in question and the type of context as well.

5.2 Limitations & Future Directions

One direction that research could expand upon in the future is examining other manipulations of familiarity. As previously stated, selecting this specific manipulation and only using this one manipulation throughout the experiments was purposeful. However, in doing so, I was unable to determine whether other manipulations of familiarity would have the same, or similar, type of impact on words and objects, in comparison to faces. That is, would a "famous" word (i.e. Mt. Rushmore) or object (i.e. the Stanley Cup) receive the same CR benefit as a non-famous word (i.e. mountain) or object (i.e. a less familiar trophy or cup), mirroring the results of my studies? Or would the pattern of results now follow that for faces? Some past evidence suggests the latter. More meaningful words have been shown to receive a reduced benefit of context (Isarida et al., 2012). Evidence has yet to be collected for famous and non-famous target objects. Further investigation is required to determine the exact influence a fame manipulation may have on object context effects.

There is, however, evidence that a fame (rather than pre-exposures) manipulation can produce the opposite pattern as what I observed in Experiments 3, 4, and 5 for target faces; Reder and colleagues (2013) found that famous faces actually receive an enhanced CR benefit. It may then be that manipulating familiarity in my thesis, and manipulating fame status, in Reder et al.,'s (2013) study are different encoding manipulations entirely.

A key factor that differentiates Reder et al.'s (2013) 'fame', from my 'pre-exposure' manipulation, is that lures on their recognition test could be selected by fame status as well. This lure fame status may be critical because, as shown in the data from Reder and colleagues (2013), famous lures had higher false alarm rates. It is possible that such a factor influenced participant behaviour and memory decisions. In my studies, I found that increasing familiarity reduced context effects. Specifically, though, this occurred because hit rate, when paired with novel contexts, improved with increasing number of pre-exposures. If then famous lures, which would be akin to 10 or 5 pre-exposure lures, induced a higher false alarm rate, there should be no impact on the benefit of context. That is, accuracy rates for familiar faces, whether paired with reinstated or novel contexts, should decrease, leaving my pattern of results unchanged. However, if the pattern induced by a fame manipulation is indeed different than that demonstrated in this thesis, even further research is required to understand what distinguishes a pre-exposures manipulation from a manipulation of fame status.

An additional step that researchers in psychology must also often wrestle with is whether the phenomenon they are dealing with can be applied to real life circumstances. Is the CR effect applicable to real life? There is research in the eyewitness memory literature that suggests it very well might be and returning to the scene of the crime can enhance eyewitness recall of a particular event (Hammond et al., 2006; Krafka & Penrod, 1985; Malpass & Devine, 1981; Smith & Vela, 1992; Wong & Read, 2011). Anecdotal experience also tells us that context is not necessary to recognize a friend or a family member, with whom we are very familiar. This anecdote is very much in line with my findings here, whereby, little to no CR benefit was observed for familiar (5 and 10 pre-exposures) faces. However, such a question has yet to be answered empirically. Until such time, future research should make an attempt to determine if the phenomenon of familiarity and the context benefit can be extrapolated from the lab to general real life scenarios.

5.3 Future Research on the Context Reinstatement Effect and Memory

As a whole, the results of my thesis suggest that it is important for researchers to consider the features and aspects of the stimuli that are chosen for any CR effect study. Throughout the experiments described in this thesis, I have demonstrated that the features of both the target and the context can largely impact the benefit of context effects.

This thesis may help to answer questions about the inconsistencies in findings across the field (see Smith & Vela, 2001 for review). Across the vast number of studies that contribute to the literature, most use different sets of stimuli for both targets and context and even use different tests of memory. Within the category of context, there are also many different arguments for features of the context that influence the CR effect, such as local (i.e. semantic context which is associated with a single or a few stimulus items) and global (i.e. the environment which is associated with many stimulus items; Dalton, 1993), interactive (i.e. when the context modifies the meaning of a word, as in the pair cold-ground) and independent (i.e. when the background environment and a word are encoded individually; Baddeley, 1982), and nonincidental and incidental (i.e. spatial and temporal contexts not obviously related to the targets on a memory test; Smith & Vela, 2001). Context also fails to boost memory when it lacks semantic meaning (i.e. an intact vs. scrambled face; Skinner & Fernandes, 2010). With so many different interpretations of the CR effect, it becomes difficult to discern when and under which circumstances the CR effect is effective and when it is not. Of course, it is beneficial to provide converging evidence of context memory benefits to aid in our understanding of the effect. However, I wish to illuminate how varied CR research can be and suggest special care be taken when discussing when and to which stimuli context effects can be applied.

In addition to variations in context, there are a number of theories that attempt to capture CR effects. I emphasized global matching models and the outshining hypothesis in this thesis. However, it should be noted that there are some global matching models that have different hypotheses for the influence of familiarity. For example, simulations of MINERVA2 (Hintzman, 1984) makes similar predictions as the outshining hypothesis. Specifically, MINERVA2 predicts a decline in CR benefit with an increase in familiarity. The discrepancy in the predictions of global matching models suggest that perhaps there is a larger discussion to be had about which theories fit under the umbrella of the global matching models.

In addition to global matching models, there are other theories that offer suggestions for the cognitive mechanisms underlying the CR effect. For example, relational binding theory (Shimamura, 2003) similarly suggest that items, when paired together, become bound. Repeated activations of such a binding stabilizes the connection. The resulting binding that is formed creates the context benefit and enhanced memory performance. The multiple-trace binding theory (Nadel & Moscovitch, 1997) also suggests that reactivation of memory representations produces different, but related neural traces. Recently bound reactivations produce additional links within the hippocampal cortex. As a result, the item becomes related to a variety of cues, thus increasing the likelihood or reactivation. In other words, various contexts are able to aid memory for the target. It is interesting to note that then, while both theories support the CR effect, each add a nuance to how context aids familiar items. For example, relational binding may offer a better explanation for memory for the butcher in his butcher shop. That is, one item (the butcher) that is repeatedly paired with one context (the butcher shop) receives a large benefit of context. Anecdotal experience with a friend, conversely, may be more in line with the predictions of the multiple-trace theory. Since friends form traces with such a large number of cues, perhaps the strength of the memory representation for the target (friend) becomes independent of context. Further research is required to test and contrast these theories to one another.

Another possible avenue for the field and the general population is to use context cues to facilitate activities and memories for the aging population. Specifically, it may be that context could exert especially beneficial influences in this population, for certain target materials, as their memory worsens. Age-related declines in associative (Bayen, Phelps, & Spaniol, 2000) and source (Hashtroudi, Johnson & Chrosniak, 1989; Spencer & Raz, 1994, 1995) memory are well documented to this point. However, there is some evidence that the CR effect may be especially beneficial for older adults (Craik & Schloerschiedt, 2011; Naveh-Benjamin & Craik, 1995; Robin & Moscovitch, 2017; Vakil, Melamed, & Even, 1996, but see Rose, Bull, & Vrij, 2003). If true, context reinstatement may offer a solution, particularly for older adults facing memory declines, or who are attempting to acclimate to new environments, such as retirement homes. Older adults, especially those with Alzheimer's or other forms of dementia, often experience disorientation when moving from their home to a retirement community. Equipping retirement homes with familiar pictures, or even furniture, may enhance the memory of older adults, enabling them to feel more comfortable and ease their transition to their new environment.

5.4 General Conclusions

Throughout the experiments described in this thesis, I have demonstrated that specific features, such as the stimulus type, familiarity within a stimulus type and anxiety-provoking level of a context, can influence the CR benefit. Importantly, the findings of these studies provide evidence that various factors and features of the items and contexts that compose item-context pairs can influence later memory decisions. The CR effect does not always apply, particularly when applied to recognition memory of faces. Indeed, familiarity of faces, in contrast to words and objects, seems to influence the benefit context can offer to memory. That is, when faces are highly familiar or encountered in specific contexts, reinstating context does not aid memory. Instead, other factors, such as the memory signal strength of the face, play a larger role and are more important in determining recognition memory decisions.

References

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 997-1010. https://doi.org/10.1037/0278-7393.25.4.997

Amir, N., Beard, C., & Bower, E. (2005). Interpretation bias and social anxiety. *Cognitive Therapy and Research*, *29*(4), 433-443. https://doi.org/10.1007/s10608-005-2834-5

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review, 89*(6), 708-729. https://doi.org/10.1037/0033-295X.89.6.708

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*(1), 1. https://doi.org/10.1037/0033-2909.133.1.1

Bayen, U. T., Phelps, M. P., & Spaniol, J. (2000). Age-related differences in the use of contextual information in recognition memory: A global matching approach. *The Journals of Gerontology: Series B, 55*(3), 131-141. https://doi.org/10.1093/geronb/55.3.P131

Bisby, J. A., & Burgess, N. (2017). Differential effects of negative emotion on memory for items and associations, and their relationship to intrusive imagery. *Current Opinion in Behavioral Sciences, 17*, 124-132. https://doi.org/10.1016/j.cobeha.2017.07.012

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49-59.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical Report C-1, The Center for Research In Psychophysiology, University of Florida.

Brown, J. M. (2003). Eyewitness memory for arousing events: Putting things into context. *Applied Cognitive Psychology, 17*(1), 93-106. https://doi.org/10.1002/acp.848

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical Psychology Review, 30*(2), 203-216. https://doi.org/10.1016/j.cpr.2009.11.003

Connor, K. M., Davidson, J. R., Churchill, L. E., Sherwood, A., Weisler, R. H., & Foa, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN): New self-rating scale. *The British Journal of Psychiatry*, *176*(4), 379-386. https://doi.org/10.1192/bjp.176.4.379

Constans, J. I., Penn, D. L., Ihen, G. H., & Hope, D. A. (1999). Interpretive biases for ambiguous stimuli in social anxiety. *Behaviour Research and Therapy*, *37*(7), 643-651. https://doi.org/10.1016/S0005-7967(98)00180-6

Craik, F. I., & Schloerscheidt, A. M. (2011). Age-related differences in recognition memory: Effects of materials and context change. *Psychology and Aging, 26*(3), 671-677. https://doi.org/10.1037/a0022203

Craik, F. I., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, *11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dalrymple, K. A., Fletcher, K., Corrow, S., das Nair, R., Barton, J. J. S., Yonas, A., & Duchaine, B. (2014). "A room full of strangers every day": The psychosocial impact of developmental prosopagnosia on children and their families. *Journal of Psychosomatic Research, 77*(2), 144-150. https://doi.org/10.1016/j.jpsychores.2014.06.001

Dalton, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory & Cognition, 21*(2), 223-234. https://doi.org/10.3758/BF03202735

Dougal, S., & Rotello, C. M. (1999). Context effects in recognition memory. *The American Journal of Psychology, 112*(2), 277-295. http://dx.doi.org/10.2307/1423354

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89*(6), 627-661. https://doi.org/10.1037/0033-295X.89.6.627

Emmerson, P. G. (1986). Effects of environmental context on recognition memory in an unusual environment. *Perceptual and Motor Skills, 63*(3), 1047-1050. https://doi.org/10.2466/pms.1986.63.3.1047

Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. https://doi.org/10.3758/BF03193146

Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition, 13*(4), 333-345. https://doi.org/10.3758/BF03202501

Fine, D. R. (2012). A life with prosopagnosia. *Cognitive Neuropsychology, 29*(5-6), 354-359. https://doi.org/10.1080/02643294.2012.736377

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*(3), 325-331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology, 71*(1), 99-104. https://doi.org/10.1111/j.2044-8295.1980.tb02735.x

Gruppuso, V., Lindsay, D. S., & Masson, M. E. J. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review, 14*(6), 1085-1089. https://doi.org/10.3758/BF03193095

Hammond, L., Wagstaff, G. F., & Cole, J. (2006). Facilitating eyewitness memory in adults and children with context reinstatement and focused meditation. *Journal of Investigative Psychology and Offender Profiling, 3*(2), 117-130. https://doi.org/10.1002/jip.47

Hanczakowski, M., Zawadzka, K., & Macken, B. (2015). Continued effects of context reinstatement in recognition. *Memory & Cognition, 43*(5), 788-797. https://doi.org/10.3758/s13421-014-0502-2

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4*(1), 106–112. https://doi.org/10.1037/0882-7974.4.1.106

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of Vision, 8*(1), 1–10. https://doi.org/10.1167/8.1.9.Introduction

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory and Cognition, 33*(1), 98–106. https://doi.org/10.3758/BF03195300

Herrera, S., Montorio, I., Cabrera, I., & Botella, J. (2017). Memory bias for threatening information related to anxiety: An updated meta-analytic review. *Journal of Cognitive Psychology, 29*(7), 832-854. https://doi.org/10.1080/20445911.2017.1319374

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*(2), 96-101. https://doi.org/10.3758/BF03202365

Hockley, W. E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1412-1429. https://doi.org/10.1037/a0013016

Holm, L., & Mäntylä, T. (2007). Memory for scenes: Refixations reflect retrieval. Memory and Cognition, 35(7), 1664–1674. https://doi.org/10.3758/BF03193500

Hou, C., & Liu, Z. (2019). The survival processing advantage of face: The memorization of the (un)trustworthy face contributes more to survival adaption. *Evolutionary Psychology, 17*(2), 1-12. https://doi.org/10.1177/1474704919839726

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology, 33*(1), 36-67. https://doi.org/10.1016/0022-2496(89)90003-5

Humphreys, M. S., & Chalmers, K. A. (2016). *Thinking about human memory.* Cambridge University Press. https://doi.org/10.1017/CBO9781316091920

Isarida, T., Isarida, T. K., & Sakai, T. (2012). Effects of study time and meaningfulness on environmental context-dependent recognition. *Memory & Cognition, 40*, 1225-1235. https://doi.org/10.3758/s13421-012-0234-0

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989a). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56*(3), 326-338. https://doi.org/10.1037/0022-3514.56.3.326

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989b). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118*(2), 115-125. https://doi.org/10.1037/0096-3445.118.2.115

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology, 49*(1), 58-69. https://doi.org/10.1037/0022-3514.49.1.58

Lee, C. M., & Fernandes, M. A. (submitted, CEP-2020-0243) Pre-exposures to faces (but not words or objects) reduces the context reinstatement benefit to memory. *Canadian Journal of Experimental Psychology*

Liebowitz, M. R. (1987). Social phobia. *Modern Problems in Pharmacopsychiatry, 22*, 141–173.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122-1135. https://doi.org/10.3758/s13428-014-0532-5

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 153-161. https://doi.org/10.1037/0278-7393.28.1.153

Malpass, R. S., & Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology, 66*(3), 343-350. https://doi.org/10.1037/0021-9010.66.3.343

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252-271. https://doi.org/10.1037/0033-295X.87.3.252

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science, 6*(2), 114-133. https://doi.org/10.1177/1745691611400234

Meade, M. E., & Fernandes, M. A. (2017). Semantic and visual relatedness of distractors impairs episodic retrieval of pictures in a divided attention paradigm. *Visual Cognition, 25*(7-8), 825-840. https://doi.org/10.1080/13506285.2017.1344341

Miles, W. R. (1930). Ocular dominance in human adults. *The Journal of General Psychology, 3*(3), 412-430. https://doi.org/10.1080/00221309.1930.9918218

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*(4), 630-633. https://doi.org/10.3758/BF03206543

Mitte, K. (2008). Memory bias for threatening information in anxiety and anxiety disorders: A meta-analytic review. *Psychological Bulletin, 134*(6), 886. https://doi.org/10.1037/a0013343

Mogg, K., McNamara, J., Powys, M., Rawlinson, H., Seiffer, A., & Bradley, B. P. (2000). Selective attention to threat: A test of two cognitive models of anxiety. *Cognition & Emotion*, 14(3), 375-399. https://doi.org/10.1080/026999300378888

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory. *Psychological Review, 104*(4), 839-862. https://doi.org/10.1037/0033-295X.104.4.839

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 882-894. https://doi.org/10.1037/0278-7393.19.4.882

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition, 22*(5), 584-590. https://doi.org/10.1037/0278-7393.21.1.158

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 158-172. https://doi.org/10.3758/BF03198397

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General, 128*(4), 403-415. https://doi.org/10.1037/0096-3445.128.4.403

Nadel, L., & Moscovitch, M. (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7*(2), 217-227. https://doi.org/10.1016/S0959-4388(97)80010-4

Naveh-Benjamin, M., & Craik, F. I. M. (1995). Memory for context and its use in item memory: Comparisons of younger and older persons. *Psychology and Aging, 10*(2), 284–293. https://doi.org/10.1037/0882-7974.10.2.284

Olsen, R. K., Sebanayagam, V., Lee, Y., Moscovitch, M., Grady, C. L., Rosenbaum, R. S., & Ryan, J. D. (2016). The relationship between eye movements and subsequent recognition: Evidence from individual differences and amnesia. *Cortex, 85*, 182–193. https://doi.org/10.1016/j.cortex.2016.10.007

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America, 109*(48), 3314–3323. https://doi.org/10.1073/pnas.1214269109

Rainis, N. (2001). Semantic contexts and face recognition. *Applied Cognitive Psychology, 15*(2), 173-186. https://doi.org/10.1002/1099-0720(200103/04)15:2<173::AID-ACP695>3.0.CO;2-Q

Reder, L. M., Victoria, L. W., Manelis, A., Oates, J. M., Dutcher, J. M., Bates, J. T., Cook, S., Aizenstein, H. J., Quinlan, J., & Gyulai, F. (2013). Why it's easier to rmemeber seeing a face we already know than one we don't: Preexisting memory representations facilitate memory formation. *Psychological Science, 24*(3), 363-372. https://doi.org/10.1177/0956797612457396

Robin, J., & Moscovitch, M. (2017). Familiar real-world spatial cues provide memory benefits in older and younger adults. *Psychology and Aging, 32*(3), 210-219. https://doi.org/10.1037/pag0000162

Rose, R. A., Bull, R., & Vrij, A. (2003). Enhancing older witnesses' identification performance: Context reinstatement is not the answer. *Canadian Journal of Police and Security Services*, *1*(3), 173.

Russo, R., Ward, G., Guerts, H., & Scheres, A. (1999). When unfamiliarity matters: Changing environmental context between study and test affects recognition memory for unfamiliar stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 488-499. https://doi.org/10.1037/0278-7393.25.2.488

Rutherford, A. (2004). Environmental context-dependent recognition memory effects: An examination of ICE model and cue-overload hypotheses. *The Quarterly Journal of Experimental Psychology, 57*(1), 107-127. https://doi.org/10.1080/02724980343000152

Shechner, T., Britton, J. C., Pérez-Edgar, K., Bar-Haim, Y., Ernst, M., Fox, N. A., Leibenluft, E., & Pine, D. S. (2012). Attention biases, anxiety, and development: Toward or away from threats or rewards? *Depression and Anxiety, 29*(4), 282-294. https://doi.org/10.1002/da.20914

Shimamura, A. P. (2003) Relational Binding Theory. *Neuropsychology of memory*, 61.

Schonbach-Medina, S., & Vakil, E. (2017). Encoding factors affecting context effects on memory: Congruency, attention and exposure time. *Psychology, 8*(3), 463-476. https://doi.org/10.4236/psych.2017.83029

Skinner, E. I., & Fernandes, M. A. (2010). Effect of study context on item recollection. *The Quarterly Journal of Experimental Psychology, 63*(7), 1318-1334. https://doi.org/10.1080/17470210903348613

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5*(5), 460-471. https://doi.org/10.1037/0278-7393.5.5.460

Smith, S. M. (1985). Environmental context and recognition memory reconsidered. *Bulletin of the Psychonomic Society, 23*(3), 173-176. https://doi.org/10.3758/BF03329818

Smith, S. M. (1986). Environmental context-dependent recognition memory using a short-term memory task for input. *Memory & Cognition, 14*(4), 347-354. https://doi.org/10.3758/BF03202513

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies, & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13-34). Oxford, England: John Wiley & Sons.

Smith, S. M. (1994). Theoretical principles of context-dependent memory. In P. E. Morris, & M. Gruneberg, *Theoretical aspects of memory* (2nd ed., pp. 168-195). London: Routledge.

Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology, 6*(2), 125-139. https://doi.org/10.1002/acp.2350060204

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*(2), 203-220. https://doi.org/10.3758/BF03196157

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*(4), 342-353. https://doi.org/10.3758/BF03197465

Spencer, W. D., & Raz, N. (1994). Memory for facts, source, and context: Can frontal lobe dysfunction explain age-related differences? *Psychology and Aging, 9*(1), 149-159. https://doi.org/10.1037/0882-7974.9.1.149

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10*(4), 527-539. https://doi.org/10.1037/0882-7974.10.4.527

Sterling, L., Dawson, G., Webb, S., Murias, M., Munson, J., Panagiotides, H., & Aylward, E. (2008). The role of face familiarity in eye tracking of faces by individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(9), 1666-1675. https://doi.org/10.1007/s10803-008-0550-1

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson, *Organization of memory* (pp. 382-404). New York: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1. https://doi.org/10.1037/h0080017

Vakil, E., McDonald, S., Allen, S. K., & Vardi-Shapiro, N. (2019). Facial expressions yielding Context-Dependent Effect: The additive contribution of eye movements. *Acta Psychologica, 192*, 138–145. https://doi.org/10.1016/j.actpsy.2018.11.008

Vakil, E., Melamed, M., & Even, N. (1996). Direct and indirect measures of contextual information: Older versus young adult subjects. *Aging, Neuropsychology, and Cognition, 3*(1), 30-36. https://doi.org/10.1080/13825589608256610

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior, 15*(5), 505-517.

Wong, C. K., & Read, J. D. (2011). Positive and negative effects of physical context reinstatement on eyewitness recall and identification. *Applied Cognitive Psychology, 25*(1), 2-11. https://doi.org/10.1002/acp.1605

Wynn, J. S., Shen, K., & Ryan, J. D. (2019). Eye movements actively reinstate spatiotemporal mnemonic content. *Vision, 3*(2), 21. https://doi.org/10.3390/vision3020021

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research, 65*(5), 445-451. https://doi.org/10.1016/j.jpsychores.2008.03.013

Yeung, R. C., & Fernandes, M. A. (2019a). Social anxiety enhances recognition of task-irrelevant threat words. *Acta Psychologica, 194*, 69-76. https://doi.org/10.1016/j.actpsy.2019.01.015

Yeung, R. C., & Fernandes, M. A. (2019b). Altered working memory capacity for social threat words in high versus low social anxiety. *Anxiety, Stress, & Coping*, *32*(5), 505-521. https://doi.org/10.1080/10615806.2019.1626838