Attention as a Mechanism for Object-Object Binding in Complex Scenes

Kacie Mennie
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Cognition and Perception Commons, and the Cognitive Psychology Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/5020

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
ATTENTION AS A MECHANISM FOR OBJECT-OBJECT BINDING IN COMPLEX SCENES

A Dissertation

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

in

The Department of Psychology

by

Kacie Mennie
B.A. University of Michigan - Dearborn, 2012
M.A. Louisiana State University, 2015
August 2019
# Table of Contents

Abstract .................................................................................................................................................. iii

Introduction ............................................................................................................................................... 1

Experiment 1. Direct Binding ................................................................................................................ 31

Experiment 2. Binding Variability ......................................................................................................... 49

General Discussion .................................................................................................................................... 60

References .................................................................................................................................................. 69

Appendix A. IRB Approval ..................................................................................................................... 77

Appendix B. A’ Analyses ....................................................................................................................... 78

Vita ............................................................................................................................................................... 80
Abstract

The current study attempted to determine whether direct binding between objects in complex scenes occurs as a function of directed attention at encoding. In Experiment 1, participants viewed objects in one of these different types contexts: unique scenes, similar scenes, or arrays with no contextual information. Critically, only half of the objects were attended for each encoding trial. Participants then completed an associative recognition task on pairs of items created from the previously studied scenes. Test pairs consisted of two attended or unattended objects, and were associated with a unique scene, a similar scene, or an array. Evidence of binding for attended objects was clear. Associative recognition was better for attended pairs, relative to unattended pairs, regardless of the type of context in which the objects were studied. Object-context binding was not observed in memory for attended object pairs, but was observed for unattended object pairs. Experiment 2 explored the extent to which binding strength between object relationships varies as a function of temporal and/or spatial proximity. The procedure for Experiment 2 was identical to Experiment 1, with the exception that all of the objects in the encoding trials were attended. There were no significant main effects or interactions of spatial and temporal distance on binding strength, as measured by associative recognition.
Introduction

The concept of binding in memory concerns the process by which perceptual and other related information about a single stimulus or more complex event are tied together into a representation. The smallest unit of visual information, the feature, can be studied in terms of its integration with the object to which it belongs, or its integration with other features to produce perception of the object (Treisman & Gelade, 1980). Binding between larger, more complex pieces of visual information can be studied in the same way. A larger, more complex binding might be defined as an object and its context, with the context made up of multiple objects. One example is remembering how you know a person (the object) from a particular party at a friend’s house (the context). One question that can be asked is how binding between two objects in a scene occurs and how this is influenced by the broader scene context.

Regardless of scale or complexity of stimuli, each type of binding can be considered a form of associative memory, or memory for relationships, the study of which is pervasive throughout visual-working memory (VWM) and long-term memory (LTM). That said, the coverage of different aspects of binding that occur within a complex scene is more robust in some areas than others, limiting the ability to draw parallels across the types of binding to inform predictions about the other. For example, there is a feature-level literature that explores whether features are represented as bound units in memory (e.g., Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Awh, Barton, & Vogel, 2007; Brady & Tenebaum, 2013; Karlsen, Baddeley, & Hitch, 2010; Zhang & Luck, 2008; Adam, Vogel, & Awh, 2017), or independently of each other (Ma, Bays, &
Husain, 2014; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004; Bays, Catalao, & Husain, 2009; Starns & Hicks, 2005; Starns & Hicks, 2008). There is also much research on scene memory that explores the relationship between a target object and its context (e.g., Murnane, Phelps, Malmberg, 1999; Lampinen, Copeland, & Neuschatz, 2001; Hollingworth, 2006a, 2006b, 2007, Nakashima & Yokosawa, 2011; Martin, 2013; Horner & Burgess, 2014; Udale et al., 2017). Missing from the literature is an examination of how two objects in a scene are related to each other, which leaves open the question of whether objects bind only to the broader scene context, as opposed to the broader context and to each other. The logical progression in the research should be to extend the findings of feature-feature binding literature to object-object binding literature, and then apply those findings to scene/context research. It could be argued that the large body of research on associative memory between two objects in LTM is evidence of this progression to binding of more complex stimuli. However, that type of research explores the relationship between two items in isolation (e.g. Murray & Kensinger, 2012; Diana, Reder, Arndt, & Park, 2006; Gronlund & Ratcliff, 1989). This research does not consider contextual factors (like scene information) that may influence the binding of two objects, paralleling the way in which single object information impacts feature relationships of that object. In essence, research on the relationship between two features, an object and its context, or two objects absent any context has not been integrated, creating gaps in the theoretical assumptions about binding of visual information.

Besides the lack of research on the topic, there are a number of broad reasons to
examine whether object-object binding occurs in complex scenes, and the mechanisms that might influence it. One is that it is of theoretical interest to understand the ways that representations of complex memories are structured (e.g., Rubin & Umanath, 2015). Another reason concerns the application of memory research to situations such as eyewitness testimony. For instance, there has been a growing interest in using formal models to predict the overall accuracy of an eyewitness (e.g., Dunn & Kirsner, 2011; Waubert de Puiseau, Ablafg, Erdfelder, & Bernstein, 2012). Understanding the conditions under which there might be a stronger or a weaker relationship between memory for some aspects of an event (say, ones that could be corroborated) and other reported elements could be a useful input to such models.

One way to approach the study of object-object binding is to consider that this type of binding is analogous to feature-feature binding. While scene-binding research and associative memory research is available, neither allows for the clearest predictions of how object bindings may behave in scenes because the former focuses on aggregate contextual influences on one object and the latter focuses on object binding without context. However, the feature binding literature does cover both of these levels of analysis. One area of research considers how two features bind to each other in a particular context (i.e., within an object). Additionally, in other areas of the feature binding literature, there is work on whether feature-feature binding occurs directly in both VWM (e.g., Luck & Vogel, 1997; Johnson, Hollingworth, & Luck, 2008; Karlsen et al., 2010; Luria & Vogel, 2011; Brady & Tenebaum, 2013; Ecker et al., 2013) and LTM (e.g., Starns & Hicks, 2005, 2008; Meiser & Bröder, 2002; Meiser & Sattler, 2007; Boywitt & Meiser, 2012).
Given that the visual system is hierarchical (Serre, 2014), a reasonable beginning assumption is that how two objects bind together in a scene could be an analogous process. For example, one major question that comes out of the feature binding literature is whether features of objects are bound to each other as well as to the object in a memory representation. The analogous question in scene memory is whether objects are bound to each other as well as to the scene context (see Fig. 1). Utilizing similar methodologies found in the feature-binding research, the current study would explore the extent to which objects in scenes bind together as a function of attention deployment at encoding. Below, I will briefly review the binding literature and then review instances under which binding strength between two objects may vary before outlining the proposed set of studies.

![Diagram](image)

Figure 1. The binding analogy. This figure is adapted from Starns and Hicks (2008) illustrating the question of feature-feature binding in objects (a) and the potential analogy of object-object binding in scenes (b).

**Binding in Memory**

While the terminology and the tasks involved in binding research can vary drastically as a function of memory system, binding of visual information is actually similar across VWM and LTM (for a review see Mennie, 2018). Critically, similar results seem to arise regardless of the complexity of the binding studied. That is, whether feature-
feature (Woodman & Vogel, 2008; Fougnie, Asplund, & Marois, 2010; Troyer, Winocur, Craik, & Moscovitch, 1999; Boywitt & Meiser, 2012) or object-scene (Henderson & Hayes, 2017; Orhan & Jacobs, 2014; Brady & Tenebaum, 2013; Nakashima & Yokosawa, 2011; Hollingworth, 2007; Martin, 2013) binding is studied, attention to the target stimuli seems to dictate whether direct binding will occur. Therefore, as a first approximation, it is reasonable to assume that object-object binding should be influenced by factors in ways that are analogous to feature-feature binding. If this is the case, then what is known about feature-feature binding can be used to make predictions about object-object binding in scenes.

Although there is a vast literature on scene and context memory that explains some aspects of object binding (for a review see Hollingworth 2006), these studies concern how objects bind to multiple objects and leave open the question of how objects in a scene are bound directly to each other. For this reason, it is sensible to extend the techniques used in studying feature binding to address this issue. Determining whether objects are capable of direct object-object binding, in addition to binding with the scene, would have important implications for how scene memory is understood. Therefore, utilizing the feature-feature binding literature in both VWM and LTM to predict under what conditions objects may bind directly to each other in LTM would be a useful starting point.

Before discussing its application and extension to the study of object-object binding, I will first review the prominent theories and a selection of empirical findings of fea-
ture-feature binding in both VWM\textsuperscript{1} and LTM. An inclusion of prominent theories is necessary given that regardless of memory system, it is unclear whether feature-feature binding occurs. Some researchers argue for direct binding (e.g., Luck & Vogel, 1997; Vogel, et al., 2001; Awh, et al., 2007; Brady & Tenebaum, 2013; Karlsen, et al., 2010; Zhang & Luck, 2008; Adam et al., 2017), while others argue for independence (e.g., Ma, Bays, & Husain, 2014; Bays, Catalao, Husain, 2009; Bays, Wu, Husain, 2011; Brady, Konkle, Alvarez, & Oliva, 2012; Founie & Alvarez, 2011). The former is defined as a direct relationship between two features, while the latter is defined as no direct relationship between two features. Support for both is found, but is contingent on which theory motivated the research. To preface, I argue that it is not an issue of whether feature-feature binding occurs, but under what circumstances. As a result, similar conditions should be considered when thinking about object-object binding in scenes.

**VWM.** Visual working-memory is a short-term capacity-limited system that is used for online visual tasks. Like working memory more generally, the capacity of VWM is about four “units” of information (e.g., Cowan 2001; Luck & Vogel, 1997; Adam, Vogel, & Awh, 2017; Brady & Tenebaum, 2013; Delevene & Bruyer, 2004). While we typically think of visual information in terms of entire integrated stimuli, it can be broken down into smaller features. A question, then, is whether the “unit” that determines the capacity of VWM is the entire object, or the individual feature. In the former, visual features are necessa-

\textsuperscript{1} A note on VWM – The stimuli used to study feature binding in VWM are often extremely simplistic such that it can be argued that the “object” being studied is actually two features. For example, this type of research often studies the location, color, orientation, or shape of an abstract object that is only comprised of two features. For the purposes of the current research, these studies will be treated as if they are examining binding between features.
ly bound to each other, creating perception of an object (e.g., Treisman & Gelade, 1980). In the latter, features are not bound to each other. Instead, they are independent with each feature remembered or forgotten irrespective of the other features that comprise an object (e.g., Ma et al., 2014; Bays et al., 2009; Bays et al., 2011; Brady et al., 2012).

**Object as the unit of VWM.** If features are not bound together in VWM, then memory for one feature will not be predictive of memory for another in the same object. However, there is a wealth of evidence stemming from object-based models of VWM that suggests otherwise. Most notably, visual information is typically perceived as an object, which is inherently a bound visual representation of the relationships among features (Awh, et al., 2007). Additionally, empirical evidence from change detection tasks suggests that features are bound to each other (Awh et al., 2007; Brady & Tenebaum, 2013; Delevenne & Bruyer, 2004; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Johnson et al., 2008). In a change detection paradigm, participants study a set of visual information (like an array of colored squares), which varies in the number and complexity of the items in the array. The set sizes range from one to six items such that each trial array is either within the capacity of VWM or exceeds the capacity of VWM. Additionally, the objects can vary in their complexity by the number of features that they possess (see Fig. 2). After the encoding phase, participants complete a test phase, in which they are presented with a similar set of stimuli and are asked whether the probed item is the same or different from what was originally presented (see Fig. 3). There is evidence of feature-feature binding when VWM capacity estimates are not limited by the number
of features for each object, but by the total number of objects presented in each array (e.g., Luck & Vogel, 1997; Vogel et al., 2001; Awh et al., 2007; Brady & Tenebaum, 2013).

Figure 2. Examples of how stimuli complexity may be utilized in a change detection paradigm. The figures in a and b are one shape with two different colors. In c, the texture of the square is more complex than a solid color. In d, the shape is more complex than a standard square.

A classic example of this paradigm is Luck and Vogel's (1997) work. Participants in their study studied objects that varied in complexity, with the presented objects varying in shape, color combinations, spatial orientation. At test, participants were asked to determine whether the probe object changed in regard to one feature or if both features changed (a conjunction). Capacity limits were estimated to be the same regardless of detection condition. That is, memory for the additional feature in the conjunction conditions did not impact memory capacity, suggesting that the number of items and not the number of features limit VWM capacity. Luck and Vogel (1997) concluded that the unit
Figure 3. An example of a change detection paradigm in which a change has occurred. First, the study display would be presented to the participant, followed by a delay interval. Then, participants would be presented with a test display and would have to determine whether the stimuli presented changed from study to test. In this case, the square and circle swapped color from study to test.

of VWM is the bound object. One reason for this could be that attention at encoding and retrieval is focused on the entire object (Luria & Vogel, 2011), creating a bound object representation. One way to test this hypothesis is to determine in what way VWM is disrupted under conditions of divided attention (DA). Attention is divided by dual tasks, or a situation in which there is a main task that must be completed in addition to a second, also difficult task. For example, participants might be asked to generate random numbers or complete math problems while attempting to remember a set of items, which divides attention between two tasks. As the main task and the secondary task increase in difficulty, performance decreases. For instance, memory for items encoded
under DA tends to be worse for items encoded under full attention (e.g., Clark-Foos, & Marsh, 2008; Fernandes, & Moskovitch, 2000; Lane, 2006). Given the robust disruptive effects of DA, it can be reasoned that if the unit of VWM is the feature, then memory for multiple features in an array should be disrupted under DA, relative to memory for a single feature studied under DA. However, if the unit of VWM is a bound representation of features, performance (as measured by memory accuracy) should be equivalent for objects containing multiple features (i.e., conjunctions) and single feature objects. In line with this notion, Johnson and colleagues (2008) have found that DA impacts memory for single features and conjunctions of features in the same way. If features were represented independently of each other, single features should have been easier to remember under conditions of divided attention. Instead, their results showed that memory was roughly equivalent for both conjunctions of features and single features, suggesting that the unit of VWM is the object and not the feature.

**Feature as the unit of VWM.** Even so, it is obvious that human memory does not operate in an all-or-none fashion, as object-based models of VWM would suggest. Aspects of objects are sometimes forgotten, while other features of an object remain intact. Resource sharing models are able to account for this phenomenon by conceiving of VWM capacity not as discrete, but as continuous, with the feature as the unit of information. In these models, capacity is determined by the amount of detail, or features, present in a visual display (e.g., Bays, et al., 2009; Bays, Wu, & Husain, 2011; Brady, et al., 2012). As more features are included in a display, more attentional resources are used. Critically, this occurs regardless of whether the features belong to one object or
many (e.g., Ma, et al., 2014). As a result, resource-sharing models predict that features are necessarily independent of each other at encoding and do not bind together at encoding to create an object. This binding process must occur at retrieval. Otherwise, bound feature representations, and not individual features, limit capacity (e.g., Fougnie & Alvarez, 2011; Brady et al., 2012; Park, Sy, Hong, & Tong, 2017).

Evidence for resource-sharing models comes from precision paradigms (Ma, et al., 2014; van den Berg et al., 2012; Wilken & Ma, 2004). It is argued that precision measures allow for a better measurement of rates of forgetting, relative to standard change detection paradigms, because of the continuous nature of the response. The study phase for precision paradigms is identical to the study phase of a change detection task. However, at test participants make a judgment using a sliding scale or wheel to match what was studied previously. This allows for a continuous measurement of accuracy, as opposed to a discrete same/different response. For example, an array of four lines in different colors and orientations may be presented at study. At test, a probed line would be at an orientation and in a color that was not used in the study phase. Participants then determine the orientation and the color, separately, of the probed line using the sliding scale or wheel. If binding occurs, precision for each feature should be related to the other feature of an object, but not the total amount of features in the studied array (Bays et al., 2009). That is, the precision of orientation and color should have dependent error rates, which are calculated as the deviation between what was reported by the participant and the correct answer. However, this dependency is not found in work that supports the feature as the unit of VWM. Instead, what is found is that re-
search is that there is no relationship between the rates of error for the two features that belong to an object.

In one example, Bays and colleagues (2009) presented arrays of colored squares to participants. At test, participants completed a precision task for color for a probed item in the array. Critically, color memory was dependent on location memory, suggesting that location is another feature of the studied objects (the squares). It was found that the precision of memory for color declined as the set size of the studied arrays increased. This decline in precision was observed even in an increase in set size from one colored square to two. They argue that if the unit of VWM capacity was the object, the increase from one to two objects at study should not have harmed precision, as both set sizes would be well within the limits of VWM capacity. Additionally, it was found that as set size and exposure time increased, the types of errors made were not random. Instead, they were centered on previously studied, non-target squares. In other words, a studied color, but the incorrect location was selected more often than a random, unstudied color. This suggests that the location and color information available in the study arrays was not stored as a bound representation, but as separate pieces of information that were incorrectly reintegrated at retrieval (Bays et al., 2009). Together, these results are characteristic of evidence that VWM is not object-based, but feature based. This research does not support a bound object representation that is either remembered or forgotten, but a grouping of feature memory representations that vary in precision. Accordingly, the total feature set size, not the total object set size, influences VWM capacity. In instances in which the entirety of an object is remembered, inde-
pendence theories argue that each feature of an object was encoded, retrieved, and re-integrated successfully (Adam et al., 2017).

**Attention determines binding.** It may be that the “unit” of VWM capacity is both the object and the feature. However, which one is used is likely influenced by the focus of attention at encoding (Brady & Tenebaum, 2013; Orhan & Jacobs, 2014). Environmental, bottom-up influences as well as goal-directed, top-down influences may impact the degree to which features are bound to each other (van Lamsweerde, Beck, & Johnson, 2016). As a result, features may look completely bound to each other or, features may look independent, resulting in different rates of forgetting (e.g., Bays et al., 2009).

The notion that the unit of VWM depends on the focus of attention comes from Brady and Tenebaum’s (2013) work that involved the creation of a Bayesian VWM model in which VWM representations are probabilistic and depend partly on task demands. They reasoned that summary statistics are always extracted from groups of visual information, even random ones. Strategic use of summary statistics as a way to quickly gather information about the visual environment necessitates a relationship among all of the items in view. When this type of relational information cannot be used, as in truly random displays, VWM capacity approaches the capacity limits of object-based models. This suggests a bound relationship based on attention that can be flexibly deployed to either object-level (the features) or scene-level (the objects) relationships. In addition to the inclusion of statistical regularities, Orhan and Jacobs (2014) go a step further and suggest that VWM models consider why feature-feature binding would occur over object binding. In particular, they state that binding relevance must be considered when de-
termining whether feature-feature binding will occur. Taken together, this work suggests that feature-feature binding occurs, but is dependent on the type of information attended to at encoding.

**LTM.** The feature binding debate in LTM is similar to the feature binding debate in VWM. However, whether features are bound together is an issue for recollection, but not familiarity. Recollection involves memory for detail and relationships between content and context (Yonelinas, 2002). Familiarity is a “feeling of knowing” with an absence of specific detail (Jacoby, 1991). An example of recollection is when you meet an acquaintance for the first time and later remember the time and location of the meeting. Familiarity, on the other hand, is when you decide that you have seen a person before without knowing the context from which you know them. Necessarily, memory for features and their relationships to each other (and the object) are present only in recollected memories (Meiser & Bröder, 2002; Meiser & Sattler, 2007, Starrs & Hicks 2005, 2008). Provided that a memory is recollected, support for either binding or independence is fundamentally similar to what is found in VWM.

Multidimensional source monitoring (MDSM) paradigms are used to study feature binding in LTM. The items used in MDSM vary on two features, which are the targets of memory later on. For example, participants can study an item (like a line drawing of a common object, or a word). Each item is presented in one of two colors (say, red and blue) and on either the left or right side of a computer screen. The stimuli are presented sequentially at study in their given color and location. At test, participants are presented again with a set of stimuli containing old and new items. This time, the stimuli are pre-
sented in an unstudied color (say, black) and in an unstudied location on the screen, say the center. Participants must decide whether the presented stimulus is old or new. If old, participants then decide what color the item was presented in and in what location during the study phase. By testing memory in this way, these tasks examine how features of an object are related to each other in LTM.

Researchers have used the presence of stochastic dependence as evidence of binding between features of an object (Meiser & Bröder, 2002, Starns & Hicks, 2005, 2008; Boywitt and Meiser, 2012, 2013; Meiser, 2014; Hicks & Starns, 2016). Stochastic dependence is a lack of independence between two features. In other words, memory for one feature is related to memory for the other. Typically, evidence for stochastic dependence arises when memory for one feature of a studied item is correlated with memory for the other related feature (e.g., Meiser & Bröder, 2002, Uncapher, Otten, & Rugg, 2006; Meiser & Sattler, 2007; Meiser, Sattler, & Weiβer, 2008; Uncapher & Rugg, 2009; Boywitt and Meiser, 2012, 2013; Meiser, 2014). In the previous example, evidence for stochastic dependence occurs if the accurate memory for color of an item cued memory for location of the same item. Using stochastic dependence as a proxy for binding between features, the research can be used to determine under what circumstances binding between features is found.

**Bound features.** There is support in LTM for an object-based representation for visual long-term memory that is comprised of bound features that create an object. A positive correlation between memorial accuracy of two features has been offered as evidence of stochastic dependence of features (Meiser & Bröder, 2002, Uncapher et al.,
For example, in their seminal study, Meiser and Bröder (2002) had participants view words that varied in their location and font. The test phase consisted of an old/new task in which participants first indicated whether the presented word was old or new. If the word was judged to be old, then the location of the word and the font size of the word were indicated. Accuracy for one source feature was positively correlated to accuracy of the other, suggesting that the features were bound to each other. A multinomial statistical model also confirmed this finding. Other work that using different methodologies has also replicated this effect (Uncapher et al., 2006; Meiser & Sattler, 2007; Meiser, Sattler, & Weißer, 2008; Uncapher & Rugg, 2009; Boywitt & Meiser, 2012, 2013; Meiser, 2014).

Independent features bound to context only. Despite evidence that features are bound directly to each other, there is also evidence that feature relationships are instead mediated through accurate memory for the entire object (including its features). Binding variability hypotheses state that while features may appear bound to each other in addition to the object, this is because of methodological issues present in studies that find this effect (e.g., Starns & Hicks, 2005, 2008; Hicks & Starns, 2016). It is argued, instead, that the relationship between two features is mediated through accurate memory of the object to which the features belong. As noted in Starns and Hicks (2005, 2008), the mediation occurs through one of two ways. The first is a methodological flaw in which the target object is present at test. In this instance, any remembered feature is potentially cued by the object and not by the other feature that belongs to the object (Starns & Hicks, 2008). Likewise, even if the object is not present for a recognition test,
it is possible that the object is being remembered with all of its features and then the appropriate feature is being selected as a response at test (Starns & Hicks, 2005). In both cases, memory for one feature is not facilitated by the other feature, but by memory for the object, to which both features are bound. To test this hypothesis, Starns and Hicks (2008) had participants complete a study phase in which line drawings of objects occurred in different screen locations and in different colors. At test, participants were either presented with a colored marker at a location and were asked to indicate whether the location was previously studied, the color was previously studied, or whether the color was studied in that location. This task removes the object and leaves only the unique color/location features as cues on which to base a memory judgment. If features were bound together, then the valid cues should have enhanced memory for the other feature. In other words, color memory should have enhanced location memory and vice versa. Contrary to this, providing a cue did not enhance memory for the other feature of the object, suggesting that the object must be present to serve as an intermediary link between two features.

Attention determines binding. While it is possible that features may bind to each other in addition to the object to which they belong, it is also conceivable that this relationship may not be direct, but mediated by object memory. As with binding in VWM, it is important to consider the role of attention in both theories. Attention at encoding is key to binding features to objects (e.g., Ahmad, Moscovitch, & Hockley, 2017; Troyer, et al., 1999; Troyer & Craik, 2000), therefore it is likely key to feature-feature binding as well. For example, Troyer et al. (1999) showed that divided attention at encoding dis-
rupted source monitoring and inhibited the memory for specific features of an item. Additionally, rates of recollection decrease when attention is divided at encoding (Yonelinas, 2002).

Given the disruptions to source memory and recollection, more generally, it is important to consider how attention deployment at encoding influences feature binding. Like the different models of VWM, binding between features in LTM may sometimes occur and other times not, depending on different levels of focus at the time of encoding. For example, when there is evidence of stochastic dependence of features, attention may be focused on the relationship between features, in addition to the relationship between features and object. On the other hand, when independence is found, attention is focused only on omnibus object-level information. This is similar to an updated multidimensional source recognition model posed by Meiser (2014). This model is adapted from the multinominal model originally developed by Meiser and Bröder (2002) and includes the possibility that features of one object may be retrieved independently of each other. The *multinomial model of joint source retrieval* highlights the reasons that some features of an object will be better remembered than others (Mesier, 2014). One reason that some features might be better remembered, or related, to each other than others is, again, attention at encoding. As noted by Boywitt and Mesier (2012) attention at encoding plays a role in the extent to which the binding of a feature to an object or another feature is observed. In their work, it is found that when attention is focused on the perceptual or featural details, feature binding is found. However, when attention is instead focused on other, non-perceptual characteristics (like animacy), feature binding is not
found. Given the critical role of attention in feature binding (either to another feature or an object), it is likely that attention deployment at encoding influences how object relationships in scenes are represented in LTM.

**Object-Object binding modulated by attention.** The representation of the relationship between two objects in a real-world scene is as yet unexplored in the literature. This leaves open the question of whether scene memory is always represented in an aggregate form, or if more detailed relationships are also incorporated into LTM. In both VWM and LTM, it is clear that attention deployment at encoding determines whether features of an object will appear independent of each other or bound together. In situations where features are found to be independent of each other, it may be that attention at encoding was not focused on feature relationships, resulting in a coarser processing of the target stimuli. In other words, attention might have been focused on potential object-object bindings and the test was not constructed to reveal evidence of such bindings (as noted in Mennie, 2018). Therefore, it is likely that binding at encoding can switch from focus on feature relationships to focus on object relationships depending on how attention is deployed. As a result, what is known about feature-feature binding can be applied to object-object binding in scenes, which is currently an unexplored area of the field.

It can be predicted that focusing attention on certain objects in a scene, but not others, will facilitate object-object binding for the objects that are the targets of focused attention. In reality, the focus of attention shifts from moment to moment and from object to object, suggesting that not all objects in a scene will be bound to each other with
equal strength. This is similar to the notion that feature-feature binding is not an all-or-none process (e.g., Brady & Tenebaum, 2013; Orhan & Jacobs, 2014; Meiser, 2014). If this is the case, it leaves open the possibility that within the subset of attended items there may be a graded strength in binding created by the “lag,” or, the space/time between attended objects. The next section discusses contiguity effects, which are essentially graded memory responses that arise because of the lag between items in list and map learning. Given the mechanism by which these effects occur, it is possible that the same variables that create contiguity effects in list and map memory may also serve as variables that influence binding strength in scene memory in a similar way.

**Contiguity Effects as Evidence for LTM Structure**

There are many variables that may affect item-to-item binding strength. In particular, given that the current work focuses on scene memory, object characteristics like relatedness, typicality, or co-occurrence of use likely impact the degree to which two items are bound to each other. However, these types of characteristics are dependent on pre-experiment context there is great variability in the extent to which two objects are deemed related or likely to co-occur together. Object relationships can also vary on one or more of these variables such that two objects may be considered related, but not likely to co-occur together. On the other hand, what makes two objects related, likely to co-occur together, or typically used together can contribute to each variable. As a result, it is difficult to construct scenes in which one of these variables can be isolated to determine how strongly two items are bound to each other. A solution is to utilize variables that still influence scene structure, but that are independent of pre-experimental context.
Two such variables are the spatial configuration of objects in scenes and the time at which objects in scenes are attended and encoded. It is conceivable that the experimentally determined spatial and temporal proximity of two objects likely influences the degree to which two objects are represented together in LTM.

Research examining the structure of LTM through list memory provides a fruitful variable: contiguity, defined by the encoding episode, and not on pre-experimental knowledge (Howard & Kahana, 2002a, 2002b). The contiguity effect is the finding that items that are studied in close proximity tend to be remembered together, relative to items that are studied farther apart (Howard & Kahana, 2002a, 2002b; Hume 1738-40). Contiguity can be defined by temporal (Kahana, 1996; Howard & Kahana, 2002a, 2002b) or spatial distance\(^2\) (e.g., McNamara, Ratcliff, & McKoon, 1984; McNamara, 1986) and is typically studied with list (Kahana, 1996) or map (McNamara et al., 1984) stimuli. As a result, this effect has not been previously discussed in terms of binding of visual information but rather as an associative memory phenomenon in which proximal information is clustered together in a long-term memory representation. But, by definition, contiguity can be considered a form of item-to-item binding in that accurate memory for an item is at least partially contingent on accurate memory for other items presented close together in either time or space. Additionally, contiguity effects are graded such that memory for distant items is weaker than memory for items presented closely together (Kahana, 1996; Howard & Kahana, 2002a, 2002b), suggesting that

---

\(^2\) Much of the research on spatial contiguity is confounded with temporal contiguity. When these variables are unconfounded, spatial contiguity is not found. This will be discussed in the Map Memory section.
bindings between items could be stronger, or weaker, depending upon when or where they were presented, relative to the just-retrieved item.

Critically, contiguity effects are evidence that item judgments can be influenced by associative information. Instead of only focusing attention on item information (like how features of an item are related), attention can also be focused on object relationship information (like the proximity of two objects). If item judgments were based solely on item memory, only the feature-feature binding within the item would be useful for accurately remembering an item. However, as is seen in much research on list (e.g., Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2009) and map (McNamara, et al., 1984; McNamara, 1986) learning, accurate judgments about single items are contingent on accurate judgments about proximal study items. Therefore, contiguity is an ideal variable to apply to the current work examining object-to-object binding. It demonstrates that item-to-item binding strength can differ with varying distances in either time or space and the existing literature can be extended to scene memory research to make predictions about structure of scenes in LTM.

In the following sections, I will first discuss the major findings of temporal contiguity effects in list learning in both free recall and recognition. In particular, I will focus on research that discusses how pre-experimental variables, like temporality, influence memory for items of lists, even when other variables (like relatedness) are accounted for (e.g., the Temporal Context Model, TCM; Kahana; 1996; Howard & Kahana, 2002a, 2002b; Sederberg et al., 2009; Polyn et al., 2009). Then I will review a selection of findings of spatial contiguity effects in map learning in both free recall and recognition. To
preview, while free recall and recognition are thought to rely on different memorial processes (Yonelinas, 2002), the contiguity effect arises in both types of tasks, allowing for research using free recall to inform the research using recognition tests.

**List Memory.** TCM is a computational model that attempts to account for contiguity effects in LTM based on context defined in terms of dynamic temporal factors. The focus of this model, and its variants (e.g., Kahana, 1996; Howard & Kahana 2002b; Sederberg et al., 2009) is on the episodically determined characteristic of temporality, irrespective of distinctiveness or semantic qualities of stimuli. That is, as items are learned in successive order, a representation of the previous items stays activated in memory while the current item is studied. This activation becomes weaker as more items are learned (Polyn et al., 2009). For example, given a list of five items (i.e., Fig. 4) the item at position one would still be activated when the item at position two is studied, causing a strong contiguity effect for that pair of items. By the time the sixth item is studied, the activation of the first item has faded. When the sixth item is retrieved later on, the first item is unlikely to be recalled in succession. Retrieval of the sixth item does not enhance activation of the first item because these items were not active together during encoding (e.g., Kahana, 1996; Howard & Kahana, 2002a; Polyn, et al., 2009).
Figure 4. An example of the level of activation of different items in a list at time points a.) Two, when only two items have been studied and b.) Six, when all items have been studied. The representation for item one is still active in memory when the second item is presented (a). However, by the time the sixth item is presented, there is no residual activation from item one (b).

Kahana (1996) has likened the contiguity effect to “jumping back in time” at retrieval to the recalled item. In doing so, items that are near that point are activated in memory and are easier to retrieve, relative to items that are farther away from that point. This model of mental time travel is a hallmark of episodic memory (e.g., Tulving, 1972) and is borne out in the data examining the effects of LTM structure through list learning in both free recall (Howard, Youker, & Venkatadass, 2008; Howard & Kahana, 1999; Kahana et al., 2009; Kahana, 2002a, 2002b) and recognition (Schwartz, Howard, Jing & Kahana, 2005). The typical encoding phase of these types of paradigms can be generalized to the following: a list, or lists, of words are presented to the participant. The time between each word and the delay between study and test can vary and can be filled with dual-task distractors, which are designed to ensure target stimuli are not held in WM\textsuperscript{3}. When participants are tested, they either complete a free recall task or a recognition task.

\textsuperscript{3} The relevance of this point will be discussed later, but is presented here for completeness.
In a free recall task, participants are asked to generate all of the words they can remember from the studied lists. In a recognition task, participants are serially presented words. For each word presented, participants decide whether the word is old or new, and may provide confidence judgments about their response. Contiguity is thought to occur when the probability of remembering items together that were presented closer together in time (e.g., items 3 and 4 in a sequence) is greater than the probability of remembering items together that are temporally distant from each other (e.g., items 3 and 7 in a sequence; Kahana 1996; Farrell & Lewandowsky, 2009).

This effect is robust in both recall and recognition. However, the contiguity effect in recall has a strong forward-lag bias that is not seen in recognition (Schwartz, et al., 2005; Kahana, Howard, & Polyn, 2008). That is, when participants complete a free-recall task, they are more likely to recall list items that follow the just retrieved items instead of items that precede the just retrieved item. It is unclear why this bias appears in free recall, but not recognition. One potential reason may be that contiguity effects that arise in free recall tasks rely on a different mechanism than contiguity effects that occur in recognition tasks. However, as noted by Schwartz et al. (2005), as well as Kahana and colleagues (2008), the temporal contiguity effect is only present in recognition tasks when participants report high confidence in their responses. High confidence in a response to neutral stimuli can be considered a proxy for recollection (Kahana et al., 2008, Selmeczy & Dobbins, 2014; Wynn, Daselaar, Kessels, & Schutter, 2019). If this is the case then contiguity effects that are found in recognition tasks that lack a forward bias is due to the same mechanisms in both free recall and recognition.
One potential criticism of TCM is that this model only accounts for lags up to six (Farrell & Lewandowsky, 2009). Farrell and Lewandowsky (2009) note that much work in support of TCM does not utilize data about stimuli that are separated by more than six intervals. That is, they argued that contiguity effects are not monotonic and that accurate memory for one item is not dependent on accuracy for a temporally close item in a graded way (Farrell & Lewandowsky, 2009). Utilizing the exact parameters of the original model, Farrell and Lewandowsky (2009) reanalyzed data from fourteen different studies that were published in support of TCM. Critically, they included data about lags greater than six. In doing so, their results indicated that the contiguity effect is not monotonic and cannot be accounted for by TCM. However, Polyn, Norman, and Kahana (2009) have suggested that nonmonotonic retrieval may be due to other stimuli characteristics, like semantic similarity. Their updated model, TCM-A (Polyn et al., 2009), shows that clustering effects based on temporal proximity still occur, but these effects can interact with memory structuring based on semantics. This is in line with earlier work by Howard & Kahana (2002b) that shows that semantic clustering in LTM is contingent on temporal proximity. Therefore, although other variables can influence memory, contiguity has its own separate impact and the results from list learning demonstrate that binding strength between items can vary based on temporal proximity.

**Map Memory.** Contiguity effects are also seen in imagery research. The evidence for such affects is based on spatial priming effects in map memory (e.g., Tollman, 1948; Kosslyn et al., 1978; McNamara, Ratcliff, & McKoon, 1984; McNamara, 1986; McNamara, Hardy, & Hirtle, 1989). Instead of clustering retrieval around a specific time point, in
this research it is found that items that are closer together in space at encoding tend to be clustered together at retrieval, resulting in faster responses for proximal items (Kosslyn et al., 1978). For example, imagine that participants are asked to learn about five locations (A, B, C, D and E) on a map. Locations A-C are spatially proximal, while locations D and E are farther away and not near each other, either (see figure 5). Participants are more likely to first remember information about locations A-C, rather than locations D-E, because they are grouped together in space. Critically, participant reaction times (RTs) are faster for spatially close items, relative to spatially distant items, suggesting that spatial information about a studied map aids retrieval (e.g., McNamara et al., 1984; McNamara, 1986; McNamara et al., 1989). Therefore, in the current set of studies, items that are spatially proximal should be more strongly bound than items that are spatially distant.

Figure 5. An example of map position that results in spatial contiguity effects. Locations a, b, and c on the map are closer together to each other than they are to either location c or d. Additionally, locations c and d are not proximal to each other or locations a, b, or c. As a result, locations a, b, and c are likely to be recalled in order and with greater accuracy because of their spatial proximity.
However, there is a major shortcoming in research on spatial priming effects: temporal and spatial proximity are often confounded (e.g., Sherman & Lim, 1991; Clayton & Habibi, 1991; McNamara, Altarriba, Bendele, Johnson, & Clayton, 1989). That is, spatially close information tends to be studied in close temporal proximity, too. When these variables are not confounded, only the temporal contiguity effect remains (e.g., Sherman & Lim, 1991; Clayton & Habibi, 1991). Interestingly, spatial priming effects will still occur for free recall tasks, but not recognition tasks (e.g., Sherman & Lim, 1991; Clayton & Habibi, 1991). One potential explanation may be that both temporal and spatial information can be used to structure LTM and which one is used is dependent on the encoding and retrieval task demands. For example, Curiel and Radvansky (1998) had participants learn map information either by pointing to objects on the map, or by naming map objects. Presumably, the former task required participants to organize information spatially, while the latter required participants to organize map information temporally. A spatial priming effect was observed for participants that completed the spatially oriented encoding task. But it was not present when participants completed the temporally oriented encoding task, suggesting that spatial organization is possible, but is dependent on task demands at encoding. McNamara and colleagues (1989) found a similar impact of retrieval task. When participants in their study completed a recognition task that required spatial information, a spatial priming effect was observed, even if the encoding task did not encourage this type of structure. Altogether, the research suggests that temporal organization of LTM may be the default, but that spatial organization of LTM can occur, if necessary.
**Contiguity and Object-Object Binding.** Temporal and spatial contiguity effects observed in list and map learning research can be considered forms of item-to-item binding. Both are ideal to determine whether binding strength variability occurs as a function of stimuli characteristics because they are variables that are directly related to the encoding episode (e.g., Howard & Kahana, 2002b). Unlike pre-experimental relatedness, temporal and spatial proximity are easily experimentally controlled. It is unclear if spatial priming effects are driven by temporal contiguity. As a result, the current study will examine both temporal and spatial proximity as potential sources for binding strength variability. It may be that temporally formatted representations are the norm for LTM, but that spatial relationships can also aid in memory structure. Utilizing these stimuli characteristics is ideal because contiguity effects are robust across a variety of stimuli, still occur in recollection-based recognition, and the variables that create these effects are not dependent on pre-experimental relationships.

**The Current Study**

While it is possible that objects are only bound to the broader scene context, a more flexible explanation is that object behavior in scenes in analogous to that of feature behavior in objects (i.e., see Fig. 1b). The first experiment in the current study examined this hypothesis and explored the extent to which objects in scenes are directly bound to each other. It was thought that objects that are the focus of attention at encoding bind directly to each other, regardless of the type of scene information in which they are studied. If direct binding of objects occurs, this should have been evident regardless of the similarity of contextual information across scenes. However, if object relationships
are mediated through overall scene memory, then interference should have accrued as contextual information increases in similarity, harming memory.

The second study examined how the factors of temporal and spatial proximity impact binding strength between attended objects in scenes. Focusing only on attended objects and encoding manipulations that are not based on a pre-experimental context, the second study helped further determine under what type of circumstances object-object binding can occur. Based on the results of list and map research, it was predicted that close proximity, defined either by time or space, should create better binding between objects resulting in a measurable response. However, as previously discussed, there is a debate in map research as to whether spatial contiguity effects are driven by temporal factors. Experiment 2 was not only able to explore whether binding strength varies based on temporal and spatial distance, evaluated contiguity effects in scene memory.
Experiment 1. Direct Binding

Experiment 1 explored the question of whether objects in scenes are bound directly to each other, or simply to the broader scene context. To do this, participants studied a series of 12 scenes depicting six typical household rooms (e.g., two kitchens; two living rooms). Scenes were divided into unique context, similar context, and array conditions. For unique scenes, the context information was different for each scene, theoretically creating a unique memory trace for each studied scene. For example, in the unique condition, participants viewed two different kitchen scenes with different background and contextual information and different target objects. In the similar condition, there were six different scenes that were presented twice. The difference from the first presentation to the second presentation was not the contextual information, but the target objects contained in the scenes. That is, in the similar condition, participants viewed two kitchen scenes with different target objects, but the contextual information in the kitchen (like the wall color and appliances) was the same in both kitchens. Finally, there was a condition in which no contextual information was present. In this “blank” condition, participants studied the target objects presented in an array on a white background. In each scene (or array⁴), there was a subset of objects to which attention was directed to through a visual cue. Participants were instructed to attend to the objects in the order that they were circled with this cue. At test, participants were presented with pairs of items. These pairings consisted either of two attended objects or two unattend-

⁴ For simplicity, the word “scene” will be used going forward to describe both the scene conditions and the array condition when there is a discussion of a procedure, prediction, etc. that applies to all study presentation types. When discussing a particular condition, I will state “unique,” “similar,” or “array” context.
ed objects. Participants determined whether the pair was old or new. Old pairs consisted of objects that were studied in the same scene/array. New pairs consisted of previously studied objects that were not presented in the same scene/array. For each pair, participants also indicated their confidence level in their old/new response.

It was hypothesized that objects would bind directly to each other if they were the focus of attention at encoding. Evidence of direct object-object binding would have been indicated by better memory for objects pairs that were attended at encoding, relative to other unattended types. This should have occurred regardless of whether the contextual information was unique. This is because a direct relationship between objects in scenes would not be dependent on scene information. In addition to direct object-object binding, it was also hypothesized that object-context binding would occur. Evidence of object-context binding would have been indicated by better memory for object pairs presented in unique contexts, relative to the similar contexts. This is because the similar contexts should have created interference among the object-context representations, making it difficult to determine which pairs went together in a particular context. Finally, it was hypothesized that both object-object binding and object-context binding on memory should interacted to boost memory performance for attended object pairs studied in unique contexts. While both types of binding should be observed across scene types, memory should be best for the attended object pairs presented in unique contexts, as these pairs will be able to benefit from both the direct object-object binding and the object-context binding to create a stronger memory trace, relative to the similar contexts or array conditions.
On the other hand, it was also possible that direct object-object binding would not occur. In this case, objects would be represented independently of each other and bound only to the broader scene context (e.g., a toaster in a particular kitchen scene). In this case, memory for an association between two objects would have been mediated by scene memory, or context. Evidence of object-object binding mediated by object-context binding would be indicated by better memory for attended object pairs, but only in the unique contexts in which the scene information is unique for each set of items. Additionally, the objects studied in the similar scenes would either not show evidence of object-object binding, or object relationships would be attenuated because the similarity across scene contexts should create interference such that it would be difficult to determine which object pairings were studied together in a given scene. While not the primary focus of this experiment, it also explored the degree to which scene information influences memory for object associations. It was possible that scene information, generally, would result in better memory for object pairs. If so, object bindings should have been better remembered in both the unique and similar contexts, relative to the array contexts. That is, object pair memory should have been better in the scene conditions overall, relative to the array conditions.

**Method**

**Participants.** Participants were recruited online from Amazon MTurk, and paid $2.00 USD for their participation. Other researchers have demonstrated the validity of data collected from MTurk and have found that there are no differences between data collected online through this system and in person (see Mason & Suri, 2011;
In order to participate, participants must have met the following requirements: they were located in North America, had completed at least 100 HITS (e.g., experiments) and had a positive rating of 95% or higher from other experimenters. These criteria were selected from previous research examining attention questions (Drew & Stothart, 2016), as well as from Amazon MTurk’s developer advice.

A power analysis for Experiment 1 was computed using G*power (Faul, Erdfelder, Buchner, & Lang, 2009). The analysis suggested that the number of total participants needed to achieve 80% power to detect a significant within-between interaction was 63. This analysis assumed: a 2 (Attention Type) x 3 (Study Type) mixed-model ANOVA with an ES $f$ of 0.40 (modest effect size $f$ calculated using $\eta^2 = 0.14$) and a type 1 error rate of 0.05. This effect size was chosen based on pilot data examining whether the lack of background at test would drastically harm memory. A total of 36 females and 54 males participated for a total of 90 participants (30 in each between-subjects condition). The mean age was 37 ($SD= 11.77$). Twelve participants (four from each between-subjects group) were removed from further analyses because their mean memory performance (measured by $d'$) was 2 SD below the mean. This left 26 participants in each condition$^5$.

$^5$ Original criterion for inclusion was based on 85% accuracy or higher on the attention questions presented at encoding, but this severely limited the sample ($n = 30$), as the average was 69%. Accuracy of at least 85% was selected as the cutoff for inclusion because these questions were presented immediately after scene viewing, so participants should have done well on these questions. However, this was not the case so it is not clear that these questions served the intended purpose. Additionally, the exclusion of these data did not change the overall results. As a result, these questions were not used for exclusionary purposes.
**Apparatus and Stimuli.** The experiment was programmed in Qualtrics and linked to Amazon MTurk for data collection.

For each condition, the stimuli consisted of 12 displays created using The Sims 4™. In the first two conditions (unique and similar) the study displays showed 12 household scenes. There were six total scene types (two of each) to ensure that familiarity and/or schema knowledge alone could not be used at retrieval to make recognition decisions. In the unique condition, each scene was different (see Fig. 6). This should have been helpful because the scene information could have been used to retrieve the relationship between the objects at retrieval (for an analogous hypothesis in feature binding see Starns & Hicks, 2008). That is, it was possible that the object relationships were not direct, but were instead mediated by memory for the objects’ relationship to the scene information. In the similar condition, there were still six scene types with 12 total scenes presented. However, the context information was identical for each scene type. For example, in the helpful condition, there were two kitchen scenes that differed not only in their target objects, but also in the contextual information, like the paint, appliances, etc. In the similar condition, the target objects still differed, but the context information remained the same. Therefore, it should not have been as useful to retrieve context-specific information to make a judgment about an object pair. Finally, there was also a condition in which there was zero scene information at study (See Fig. 6). In this blank scene condition, the target objects were presented without scene information. This condition serves to help answer the question of whether scene information aids memory for associations the way it does for item memory. In each display, there were
12 objects that could serve as target objects for later retrieval. Some target objects across scenes were similar, but each was unique to the scene, and the study as a whole. For example, it was possible that two telephones were used as target objects in two different scenes. However, each telephone was distinctive.

Critically, attention was directed to seven of the objects in each scene through a visual cue. This cue was a circle around the object to which participants were supposed to attend. Attended objects appeared on the same half of the screen or along the same scan path with no other unattended objects along that path to limit potential fixations to the objects that were not supposed to be attended at encoding. Objects were counterbalanced such that they were used equally often between attention conditions. At test, participants saw 72 object pairs. The objects were presented side-by-side on a blank, white background. This should have eliminated any overt contextual information at retrieval. Object pairings consisted of two attended objects, or two unattended objects. Half of the object pairs were “old” (seen previously in the same scene) and the other half were “new” (seen previously in different scenes). New pairs consisted of a recombination of previously studied target objects, which prevented participants from making judgments on the basis of familiarity alone. New pairings were created from the same scene type so that schema information could not be used to make a correct “new” decision. For example, new pairs for the kitchen scenes only contained objects that were studied in either of the two kitchen scenes (See Fig. 7).
Figure 6. Examples of each type of scene. In the unique condition (a), participants saw two of the same scene types, but the contextual information was different. In the similar condition (b), participants still saw two scene types with different objects, but the contextual information was the same. In the object arrays (c), participants saw arrays of target objects. The objects were the target objects from the unique and similar scenes. These objects were in their locations from the other scenes, but there was no background information to anchor the items to their location.
Figure 7. Examples of study scenes and test pairs generated from the study scenes. In figures 7a and 7b attended objects are circled. In figure 7c, examples of old and new item pairs created from the scenes in figures 7a and 7b.
Design. The experiment was a 2 x 3 mixed-model factorial design. The within-subjects factor was attention at encoding to target object pairs (attended vs. unattended). The between-subjects factor was the study display type (unique vs. similar vs. array).

The dependent measures of memory performance included discrimination, as indicated by $d'$, and $A'$. Both $d'$ and $A'$ are measures of sensitivity in signal detection theory (SDT) that are not influenced by response bias (discussed later). High sensitivity, as indicated by a large $d'$, is an indicator of good discriminability between old stimuli and lure stimuli. Low sensitivity, as indicated by a small $d'$, is an indicator of poor discriminability between old stimuli and lure stimuli (e.g., Macmillan & Creelman, 2005). A $d'$ value of 0 or below is indicative of chance or below-chance performance. $d'$ is calculated from the $z$-scores of the correct responses to old items (i.e., “hits”) and the incorrect responses to new items (i.e., “false alarms”). $A'$ is a non-parametric measure that is based directly on the amount of hits and false alarms that are made throughout the course of the experiment. It is particularly useful for situations in which the overall hit or false alarm rates are relatively low. Larger $A'$ values indicate better discriminability, with a value of 0.5 indicating chance discriminability and a value of 1 indicating perfect discriminability. Unlike $d'$, $A'$ cannot be negative, and is bounded by the values of 0.5 and 1.0, but should still show the same pattern of results as $d'$ (Macmillan & Creelman, 2005).

\[ d' = z(\text{Hits}) - z(\text{FA}) \]
\[ A' = \frac{1}{2} + \left[ \frac{((\text{Hits} - \text{FA})(1 + \text{Hits} - \text{FA}))}{(4\text{Hits}(1 - \text{FA}))} \right] \text{ if } \text{Hits} \geq \text{FA} \]
\[ A' = \frac{1}{2} - \left[ \frac{((\text{FA} - \text{Hits})(1 + \text{FA} - \text{Hits}))}{(4\text{FA}(1 - \text{Hits}))} \right] \text{ if } \text{Hits} \leq \text{FA} \]
In addition to memory performance, response bias, which is a participant preference for one response or the other for reasons other than their discriminability or sensitivity. Participants can express either a liberal or conservative bias. In the former, participants state that most test items are old. In the latter, participants state that most test items are new. This means that participants that are liberally biased will maximize their amount of hits, but will also false alarm to many lures. On the other hand, participants that are conservative in their responses will show fewer hits, but also fewer false alarms. Participants without a bias should correctly respond to old and new items on the basis of discriminability alone. Response bias was measured independently of discriminability using \( c^8 \). Like \( d' \), \( c \) is calculated using the z-scores of the hits and false alarms (Macmillan & Creelman, 2005). Negative scores indicate a liberal bias, while positive scores indicate a conservative bias.

**Procedure.** Participants were provided with informed consent prior to starting the experiment. Instructions were provided on screen. Depending on the condition, participants were instructed that they would be presented with either a number of common household scenes, or arrays of common items. They were instructed to study each display and to attend to the objects in the order in which they were circled. Participants were also informed that they should be prepared for a question on one of the items in each display. This was a multiple-choice question concerning the physical characteristics for the item. For example, participants might have been asked “What color was the vase on the desk?” Then they would select their answer from a list of provided colors.

\[ c^8 = -1/2[z(\text{Hits}) + z(\text{FA})] \]
This question was intended to serve as a check to ensure that participants were paying attention to the study information. Therefore, the object that served as the attention check object at encoding was not be included in later analyses.

Each study trial consisted of one scene that remained on the screen for the entirety of the encoding time (approximately 12s). During the encoding phase, seven objects were circled with a yellow circle in at a rate of 1s per object (one served as an object for the attention check and was not used in further analyses). There was a 300ms inter-stimulus interval (ISI) between the offset of one circle and the onset of the next. This was done to allow some spreading of attention, but not so much that attention drifted to intentionally unattended target objects. After the offset of the last circle, participants answered a multiple-choice question concerning one of the circled (e.g., attended) items. Once participants answered this question, the next trial started with a different scene and the procedure repeated. The order of the scenes was semi-random such that a random order of the scenes was set into Qualtrics for each condition.

After the encoding phase, participants completed an associative recognition task. Pairs of objects were presented side-by-side on a blank background for each test trial. Presenting the pairs on a blank background ensured that only the object information could be used from the test phase to make a judgment about the pair. For each trial, participants decided whether the objects presented were previously presented in the

---

9 The scenes were not presented in a different random order each time due to the constraints of Qualtrics. It is not possible to yoke two questions together such that participants could view the scene and then answer the attached attention question. If the study scene and the attention question were embedded in the same question, then participants would have been able to see the attention question prior to seeing the scene, which would have been undesirable.
same scene (OLD), or in different scenes (NEW). After completion of the entire experiment, participants answered a set of demographic questions and were debriefed on the nature of the experiment. On average, the entire experiment took approximately 30 minutes to complete.

Results$^{10}$

Proportions of hits and false alarms were calculated. Any proportion of 0 or 1 was adjusted according to the formula in Macmillan and Creelman (2005). This adjustment prevented infinite values from occurring so that d’ could be calculated for each condition. These d’ scores were submitted to a 2 x 3 mixed-factor analysis of variance (ANOVA) with test-pair type (attended vs. unattended) as the within-subjects factor and scene context type (unique vs. similar vs. array) as the between-subjects factor.

The array-type context was intended to serve as a baseline to measure the effect of attention on binding. That is, it was expected that object-object binding should be found in the arrays when those objects were attended at encoding. However, it was thought that memory for binding between objects presented in the arrays would be equivalent to, or worse than, the unique scene. It was expected that the scenes containing unique contextual information at encoding would serve as useful cues at retrieval for attended objects, while memory for object pairs in the similar scenes should be subject to interference from the similar contexts, resulting in poor memory. However, there was no effect of context type independently of attention, $F(2, 87) = 0.55, p = 0.58$.

$^{10}$ A’ was also calculated and followed the same pattern of results as d’. See appendix a for these analyses.
If direct object-object binding occurred, memory should be better for the pairs of objects that were attended at encoding than for unattended pairs. Additionally, if object-object binding was direct and not mediated by scene memory, this effect should have occurred regardless of the type of context. This hypothesis was confirmed, $F(1, 87) = 85.55, p<0.0001, \eta^2_p = 0.50$. Pairs that were attended were recognized more often ($M = 0.55, SE = 0.08$) than pairs that were not attended at encoding, which were recognized at rates below chance (as indicated by the negative $d'$; $M = -0.50, SE = 0.10$; See Fig. 8).

![Figure 8](image)

Figure 8. Main effect of attention on memory for object pairs.

It was predicted that associative recognition performance should have been better for attended pairs studied in unique contexts. While the attention x context type interaction was significant, it was not in the anticipated direction, $F(2,87) = 3.85, p<0.05, \eta^2_p = 0.08$. Follow-up tests revealed that context type did not significantly impact attended pairs, $F(2, 89) = 0.60, p = 0.55$. However, there was a marginally significant differ-
ence in memory performance for pairs that were not attended at encoding\(^\text{11}\) across the different context types, \(F(2, 87) = 2.6, p<0.08\). Multiple comparisons (Bonferroni) revealed that, across all context types, participants performed below chance when determining whether the unattended pairs were old or new. However, unattended object pairs in the array contexts \((M = -0.71, SE = 0.17)\) were less accurately recognized than in the unique contexts \((M = -0.17, SE = 0.17)\). So, instead of an extra cue that reduces interference and aids in attended object memory, the unique contexts appear to aid unattended object pair memory.

![Graph](image)

Figure 9. Pair type x study context type interaction.

**Response Bias.** It is possible that a bias to respond in a particular way could have occurred, which may have influenced the pattern of the results. There were no \(a\) t-tests done to determine whether the recognition for unattended pairs was significantly different from chance. Recognition for unattended pairs in unique contexts was not significantly different from chance \((p = 0.30)\). Recognition for unattended pairs in similar contexts and arrays was significantly below chance \((p< 0.01)\).
priori predictions about a criterion shift in which participants would respond either more liberally, or more conservatively, based on the manipulations to the independent variables. To determine whether participants responded in a significantly biased way, the data were grouped by context type and one-sample t-tests were conducted on c, the response bias measure, for the unattended and attended object pairs. If the mean c for the different types of object pairs is significantly different from chance, then participants in that group will have responded in a biased way. Negative mean c scores indicate a liberal bias, or a tendency to call more test trials old than new. Positive mean c scores indicate a conservative bias, or a tendency to call more test trials new than old. There was a liberal bias present in the similar context condition, across both attended and unattended object pairs. There was a liberal bias present in the unique context condition, but only in responses for the unattended object pairs. Finally, there was not a response bias present for either object pair type in the array condition (see Table 1). This means that participants that studied the objects in contexts in which a background was present had a bias to respond that items were old. This would suggest that participants may have felt that the objects presented in the context of scenery were more familiar, regardless of whether that contextual information was unique or similar.

Confidence. In addition to their recognition responses, participants rated their confidence in their responses on a 1-7 scale, with 1 indicating that they were guessing and 7 indicating that they were very sure in their response. There were no a priori predictions about confidence measures, but given the direction of the significant interaction, it was thought that confidence levels might parallel the results. That is, average memory
was significantly better for unattended object pairs that were studied in a unique context, relative to those that were studied in array contexts. This is due to, in part, to a greater number of correct rejections (i.e., correctly identifying a pair as “new”) in the unique contexts. More correct rejections than false alarms for unique context, lure pairs could be due to recollection of scene information from each of the unique contexts associated with each individual item in the lure pair. This would allow participants to correctly reject these lures more often and suggest that object-context binding is occurring for the unattended object pairs. One piece of evidence for this would be greater confidence in correct rejections to lures presented in unattended-unique context pairs, relative to the unattended-array context pairs. An independent t-test revealed that this was the case. Participants that correctly rejected lure object pairs in the unique context condition were more confident ($M = 2.5$, $SE = 0.17$) in their judgment, relative to the same judgment in the array condition ($M = 2.0$, $SE = 0.15$), $t(58) = 2.03$, $p < 0.05$.

Table 1

Response bias and performance measured by hits and false alarms in the different context types across attended and unattended object pairs. Standard error in parentheses.

* Significant at 0.05 ** Significant at 0.01 *** Significant at 0.001

<table>
<thead>
<tr>
<th>Context</th>
<th>Attention</th>
<th>t-value</th>
<th>Mean C</th>
<th>Mean Hits</th>
<th>Mean FA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique</td>
<td>Attend</td>
<td>-1.96</td>
<td>-0.12 (0.06)</td>
<td>0.62 (0.03)</td>
<td>0.45 (0.03)</td>
</tr>
<tr>
<td></td>
<td>Unattended</td>
<td>-2.74**</td>
<td>-0.13 (0.05)</td>
<td>0.51 (0.03)</td>
<td>0.57 (0.03)</td>
</tr>
<tr>
<td>Similar</td>
<td>Attend</td>
<td>-3.79***</td>
<td>-0.14 (0.04)</td>
<td>0.65 (0.03)</td>
<td>0.44 (0.03)</td>
</tr>
<tr>
<td></td>
<td>Unattended</td>
<td>-2.17*</td>
<td>-0.07 (0.03)</td>
<td>0.43 (0.03)</td>
<td>0.62 (0.03)</td>
</tr>
<tr>
<td>Array</td>
<td>Attend</td>
<td>0.68</td>
<td>0.04 (0.05)</td>
<td>0.60 (0.03)</td>
<td>0.39 (0.03)</td>
</tr>
<tr>
<td></td>
<td>Unattended</td>
<td>-0.76</td>
<td>-0.03 (0.04)</td>
<td>0.39 (0.03)</td>
<td>0.63 (0.03)</td>
</tr>
</tbody>
</table>
Discussion

Overall, Experiment 1 provided evidence of direct object-object binding in scenes, regardless of the type of information conveyed in the broader context. Participants recognized objects attended at encoding more accurately than objects that were not attended. However, performance did not benefit from a unique context at encoding. One potential reason for this outcome could be that the amount of clutter in the scenes made it difficult to remember the objects, generally. This possibility is supported by the finding of good memory (measured by d’) for attended object pairs in arrays ($M = 1.7$) relative to object pairs in unique ($M = 0.82$) and similar ($M = 0.37$) scenes. Additionally, there was a liberal response bias present in the Unique and Similar contexts for both attended and unattended object pairs, but there was no bias present in the Array context. It is possible that the additional contextual information present in the Unique and Similar scenes may have created stronger feelings of familiarity, causing participants to respond that more items were old, reducing overall performance. Another possibility is that conditions that create low discriminability might cause a shift in criterion resulting in a liberal bias.

There was also a significant interaction between attention at encoding and the context type. Instead of being driven by memory for the attended object pairs, this interaction was driven by differences in memory for the unattended object pairs. It was expected that memory for objects that were not purposely attended to at encoding would be poor, which is supported. However, memory for unattended object pairs was significantly worse in some contexts than others, suggesting that the different contexts are
impacting memory, but not for the attended objects. Follow-up analyses suggested that
the objects presented in the Unique contexts ($M = -0.17, SE = 0.17$) are not recollected
as poorly as unattended objects in the Arrays ($M = -0.71, SE = 0.17$). Memory for unat-
tended object pairs in the Similar contexts was not significantly different from memory in
either the Unique contexts nor the Array contexts. This interaction would suggest that,
while the object relationships to context is overshadowed by the relationship between
attended objects, it can play a role in impacting memory for objects that were not well
attended at encoding.

Overall, Experiment 1 suggests that when objects are attended at encoding, they
can bind directly to each other in memory. It is unclear whether this object-object bind-
ing is influenced by object-context binding. However, the significant interaction would
suggest that object-context binding occurs, but is likely overshadowed by overt attention
to objects in scenes. If object-context binding were not occurring, it seems unlikely that
memory for the unattended object pairs would have benefitted from being studied in
unique contexts, as seems to be the case. This is in line with a host of other research
that shows that objects do bind to their context. However, this is one of the first studies
to examine how the relationship between two objects is represented in LTM separately
from broader contextual scene information.
Experiment 2. Binding Variability

The results of Experiment 1 support the notion that direct object-object binding in scenes is possible for objects that are well attended at encoding. Building off of this finding, Experiment 2 tested the hypothesis that object-object binding is variable, such that the strength of a relationship between two objects can change as a function of temporal or spatial proximity. The manipulations in Experiment 2 were not included in the methodology of Experiment 1 because these manipulations affect only attended objects. According to research in list (Kahana, 1996; Kahana 2002a, 2002b; Kahana et al., 2008; Sederberg et al., 2008; Polyn & Kahana, 2002; Farell & Lewandowsky, 2009; Polyn et al., 2009; Howard et al., 2008) and map (McNamara, et al., 1984; McNamara, 1986; McNamara et al., 1989) learning, items that are studied close in either time or space tend to be clustered together in memory. As a result, a contiguity effect arises such that memory for one item is related to memory for an item that is close to it, either temporally or spatially. Critically, the relationship is monotonic in that the binding between two items in a list or on a map decreases as distance increases. That is, items that are either spatially or temporally distant do not serve as good cues for each other. Therefore, it can be argued that they are not strongly bound together in LTM. Experiment 2 used what is known about temporal and spatial contiguity effects to help determine whether object-object binding strength in scenes varies based on stimuli proximity characteristics. While not the main goal of Experiment 2, the methodology was also able to examine whether spatial contiguity effects are confounded with temporal distance effects in
scene memory, which is sometimes observed list memory (e.g., Sherman & Lim, 1991; Clayton & Habibi, 1991; McNamara, et. al, 1989).

**Method**

**Participants.** Participants were recruited through Amazon MTurk. As in Experiment 1, participants were paid $2.00 USD for their participation. The same parameters for participation outlined in Experiment 1 also applied to Experiment 2. Additionally, participants that participated in Experiment 1 could not participate in Experiment 2. A power analysis for Experiment 2 was computed using G*power (Faul, Erdfelder, Buchner, & Lang, 2009). The analysis suggested that the total number of participants need to achieve 80% power to detect a significant interaction was 92. This analysis assumed: a 2 (Spatial Proximity) x 2 (Temporal Proximity) x 2 (Context type) mixed-model factorial ANOVA with an ES f of 0.10 (small effect size) and a type 1 error rate of 0.05. The effect size used to determine the sample size changed from a moderate effect size in Experiment 1, to a small effect size in Experiment 2. This is because low power and effect sizes were observed in Experiment 1 and I wanted to ensure that I did not under-sample and potentially miss a significant effect. A total of 119 participants participated. Two were removed from further analyses because their performance on the task was two SD below the mean, leaving a total of 117 participants. Out of this sample, 48 participants were female, 69 were male, and the mean age was 37.

**Apparatus and Stimuli.** Experiment 2 was programmed in Qualtrics and linked to Amazon MTurk for data collection.
As in Experiment 1, the stimuli were constructed using *The Sims 4*. Twelve new Unique scenes were constructed in order to reduce potential issues from the amount of clutter in the scenes from Experiment 1. Like Experiment 1, the Unique scenes were divided into six categories, with two scenes per category. There were two of each of the following scenes: kitchen, living room, dining room, bedroom, study space, and children’s bedroom. The Unique scenes brightness was reduced to 20% of the original brightness so that the objects that were studied in the scenes (discussed later) were more salient. There was also an Array context in which there was zero background information present and the objects were instead presented on a blank, white background. While memory for objects in the Arrays was best in Experiment 1, simplified versions of the Unique contexts were included to determine whether spatial and temporal effects were impacted by context.

The generated objects were comprised of household items, but were not necessarily scene-specific (as in Experiment 1). This was done to ensure that the objects used in Experiment 2 were clear and identifiable. Objects were counterbalanced across the scenes so that each one appeared in each of the categories at least once. This eliminated any novelty due to a lack of a relationship between the items and the scenes. For example, a kitchen item was equally likely to appear in a kitchen scene as it was to appear in a living room scene. Instead of being embedded within the scenes, the objects were arranged in a circle and were overlaid onto the Unique scenes and the white backgrounds (to construct the Array condition). All of the objects were approximately
2000mm tall or wide, depending on their orientation\textsuperscript{12}. The objects were spaced along the circle so that they were approximately 45 degrees (arc length = 302 pixels) apart (See Fig. 10). During the study phase, participant attention was directed to each object using a red circle as a visual cue.

![Image](image.png)

Figure 10. Examples of the study phase stimuli. Objects were arranged in a circle either on top of a unique background (a), or were arranged on a white background (b). Participants were assigned to either the Unique background condition or the Array condition.

A total of 32 old object pairs were composed of objects that appeared together in the scenes previously. There were four types of old test pairs that varied in their temporal and spatial distance: Spatially Close – Temporally Close, Spatially Far – Temporally Close, Spatially Close – Temporally Far, Spatially Far – Temporally Far. Objects were categorized as temporally and spatially close or far. Additionally, an item in any given scene or array could only be used in one of these four conditions. Items that were considered temporally close were attended to immediately before or after each other.

\textsuperscript{12} The object was resized according to its orientation. That is, if the object was wider (or longer) than it was tall, it was resized on the basis of its width. If the object was taller than it was wide, it was resized on the basis of its height. This provided the most reasonable size and clearest resolution for each of the objects used in Experiment 2.
Items that were temporally far were not attended back-to-back, and were separated by at least four other objects. Objects that were spatially close were directly adjacent to each other. Objects that were spatially distant were on the opposite side of the circle (a distance of approximately 400 pixels). There was a total of eight test pairs of each type.

A total of 16 lure pairs were also created. Unlike in Experiment 1, these lures were completely novel pairs. Lure pairs consisted of unstudied objects given the nature of the study. It is impossible to create lures that are a recombination of objects from different scenes and meet the requirements of spatially/temporally close or far. The objects in the lure pairs were not studied together, so there is no way that they can fit into the different temporal and spatial distance pair types in order to create a balanced number of lures. Both old and new object pairs were presented side-by-side on a blank, white background to avoid physical reinstatement of location information that could serve as a cue for memory.

**Design.** The experiment was 2 x 2 x 2 within-subjects factorial model. The within-subjects independent variables were spatial proximity (near vs. far), and temporal proximity (near vs. far). The between-subjects independent variable was the context type (Unique context vs. Array/no context).

Temporal proximity was manipulated by the lag between items as they were attended, while spatial proximity was manipulated using the physical distance between the two objects. Attention was directed using a red circle, which circled each object.

The main dependent measure was memory accuracy, or hit rate, as indicated by the proportion of correct responses to old pairs and confidence in response. While d’
was used to measure memory in Experiment 1, it cannot be uniquely calculated for the within-subjects groups because the false alarm (FA) rate that would be used to calculate d’ would be the same for each within-subjects group because of the way in which the lure pairs were constructed. Participants also rated their confidence in their responses. Confidence was rated on a 1-7 scale, with a rating of one meaning “not at all confident,” and a rating of seven meaning “extremely confident.”

**Procedure.** The procedure for Experiment 2 was almost identical to Experiment 1. The only differences were the removal of the attention questions during the encoding phase, and the number of test items was reduced from 72 to 48. The number of test items was decreased to accommodate the 2 x 2 x 2 design and to help boost performance, as memory was quite low in Experiment 1.

**Results**

It was hypothesized that object-object binding would vary within the attended object pairs as a function of lag, or distance. In particular, associative recognition for pairs that were presented close together in time should have been better than for pairs that studied with longer lags between them. This would be consistent with research on temporal (e.g., Schwartz et al., 2005; Kahana, 1996) contiguity effects. Additionally, it was hypothesized that spatial closeness would also contribute to stronger binding, as indicated by better associative recognition for objects that were not only studied close in time, but were also spatially proximal. It was unclear whether spatial effects on object-

---

13 It was possible that the FA rate could vary across context type (the between-subjects variable). However, an independent t-test revealed that the FA rate was not significantly different across conditions, ($M_{Unique} = 0.47$, $SE_{Unique} = 0.02$; $M_{Array} = 0.45$, $SE_{Array} = 0.02$).
object binding would emerge independent of temporal effects, as some research in map learning shows that spatial effects are not typically found without significant temporal effects (McNamara et al., 1989; Clayton and Habibi, 1991; Curiel & Radvansky, 1998).

Proportions of hits were submitted to a $2 \times 2 \times 2$ factorial analysis of variance (ANOVA) with temporal proximity (near vs. far) and spatial proximity (near vs. far) as within-subjects factors and Context type (unique vs. array) as the between-subjects factor. These hypotheses were not supported by the data (all analyses $p > 0.05$; see Table 2). As can be seen in Table 3, associative recognition was equivalent across contexts types, temporal distance, and spatial distance. Given the results of Experiment 2, it appears that object-object binding does not change as a function of either temporal or spatial distance. Instead, it appears that objects are bound together as long as they are attended at encoding.

Table 2.

<table>
<thead>
<tr>
<th>EFFECT</th>
<th>RESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Distance</td>
<td>$F(1, 115) = 0.02, p = 0.89, \eta^2_p = 0.00$</td>
</tr>
<tr>
<td>Spatial Distance x Context</td>
<td>$F(1, 115) = 3.39, p = 0.07, \eta^2_p = 0.03$</td>
</tr>
<tr>
<td>Temporal Distance</td>
<td>$F(1, 115) = 0.07, p = 0.79, \eta^2_p = 0.00$</td>
</tr>
<tr>
<td>Temporal Distance x Context</td>
<td>$F(1, 115) = 0.75, p = 0.39, \eta^2_p = 0.01$</td>
</tr>
<tr>
<td>Spatial Distance x Temporal Distance</td>
<td>$F(1, 115) = 0.69, p = 0.41, \eta^2_p = 0.01$</td>
</tr>
<tr>
<td>Spatial Distance x Temporal Distance x Context</td>
<td>$F(1, 115) = 0.33, p = 0.57, \eta^2_p = 0.00$</td>
</tr>
</tbody>
</table>

That said, the interaction between spatial distance and context type approaches conventional significance (see Table 2). Follow up tests showed that associative
recognition for pairs that were spatially close together was better in a unique context \((M = 0.56, \ SE = 0.02)\) than in an array \((M = 0.48, \ SE = 0.02)\), \(t(115) = -2.57, \ p < 0.01\). As can be seen in Figure 11, when object pairs are close together and studied in a unique context, they are remembered better relative to those close object pairs that were studied without context (i.e., the array condition).

Table 3.

<table>
<thead>
<tr>
<th>Context</th>
<th>Spatial Distance</th>
<th>Temporal Distance</th>
<th>Mean Hits (\text{(proportion)})</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Array</td>
<td>Close</td>
<td>Close</td>
<td>0.49</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Close</td>
<td>Far</td>
<td>0.47</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>Close</td>
<td>0.50</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>Far</td>
<td>0.53</td>
<td>0.03</td>
</tr>
<tr>
<td>Unique</td>
<td>Close</td>
<td>Close</td>
<td>0.58</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Close</td>
<td>Far</td>
<td>0.55</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>Close</td>
<td>0.54</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Far</td>
<td>Far</td>
<td>0.52</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Figure 11. Trending spatial distance x context type interaction.

One potential reason for these results could be the fact that the majority of the data was collected in close succession. However, nine participants in the array be-
tween-subjects group were collected later than the rest. To determine whether performance from this later group of participants significantly differed from the earlier group of participants, a random selection of nine participants was taken from the earlier array group and compared to this later group. The proportion of overall hits from the earlier collected data was compared to the proportion of hits from the later collected data. An independent t-test revealed that these groups were not significantly different from each other ($p = 0.07$).

**Discussion**

Experiment 2 tested the hypothesis that object-object binding is variable, such that the strength of a relationship between two objects can change as a function of temporal or spatial proximity. It was thought that items that were studied close together, either in time or space, would be clustered together in memory. Therefore, close item pairs would be better remembered than distant item pairs, which would be indicative of better binding between the two items. Another question this design addressed is the extent to which spatial proximity causally impacts the clustering of memory representations.

Due to a technical error, a subset of nine participants were collected later than the bulk of the main data set. When the entire dataset was analyzed, there were no significant main effects or interactions. But, the results suggest a non-significant trend toward an interaction, such that pairs that were spatially proximal were better remembered than when they were spatially distant when these objects were presented with a background (i.e., context). Interestingly, there were no temporal contiguity effects. One
potential reason could be that the items that were categorized as “temporally distant” were sometimes the first and last attended items. It is possible that primacy and recency effects for each scene study block overshadowed this effect. However, as Kahana (1996) notes, clustering in memory around temporal proximity tends to occur above and beyond primacy and recency effects. There were also no significant interactions of temporal and spatial proximity. It was thought that the closeness of both encoding time and spatial location would be a boost to associative memory, as the objects would have been studied in quick succession and right next to each other, allowing almost an additional encoding of the spatially close item that was just previously studied.

In sum, the results of Experiment 2 provide, at best, weak evidence that object-object binding in scenes varies in strength as a function of spatial distance, and no evidence that it varies with temporal distance. In fact, Experiment 2 provides weak evidence that binding occurs between objects at all. Participants performed just above chance in the majority of conditions. Only the condition in which objects were spatially close and in the unique scenes provides stronger evidence of binding, more broadly, as participants correctly recognized pairs in this condition 58% of the time. Given the results of Experiment 2, it appears that objects might bind directly to each other as long as they are attended at encoding (as in Experiment 1) and that this effect is not contingent on spatial or temporal factors.

It is possible that binding occurs, and that binding strength between attended objects does vary, but the current research design was not sensitive enough to detect either. This issue could be remedied in a few ways. First, the object pairings could be
made to be more continuous. The object pairs in the current experiment were designed in a dichotomous manner such that item pairs that were close together were immediately adjacent to each other, while distal pairs were separated by at least four other items. A more graded pairing might be able to better detect differences in binding variability.

Second, attention could be more strongly manipulated. While attention was directed to each item in a particular order, participants may not have attended to the objects in the way that was intended. Instead, participants may have elected to use a more diffuse attentional deployment. Or, they may have simply focused on each individual item, but without regard to the order in which they were cued. To remedy this, attention could be directed to only a subset of items in the scenes (as in Experiment 1), which should encourage participants to study the objects that are cued. These changes could allow for more sensitive analyses of variability, such as multiple regression. Using multiple regression would allow an examination of how binding changes between objects as they become closer/farther apart.
General Discussion

Despite the extensive research conducted in the realm of visual memory, both VWM and LTM, the conclusions that have been reached about the structure of visual memory have been hampered by a lack of integration across the different areas of research. Particularly, research on the structure of the relationships among objects in complex scenes is relatively absent, despite the wealth of information available on object binding (e.g., Treisman & Gelade, 1980), context memory (e.g., Murnane et al., 1999), and associative memory (e.g., Gronlund & Ratcliff, 1989), more broadly. Understanding the nature of the relationships among event components is not only of theoretical interest (e.g., Rubin & Umanath, 2015), but of practical interest as well. As mentioned previously, there is interest in using formal models to predict accuracy of eyewitness testimony (e.g., Dunn & Kirsner, 2011; Waubert de Puiseau, Ablafg, Erdfelder, & Bernstein, 2012), so it is important to understand the conditions under which the relationships between memory components might change (either become stronger or weaker) so that this information could be used as input to such models.

In order to begin to answer the question of how objects and their relationships to each other (and the scene) are represented in memory, I considered the hierarchical nature of the visual system (Serre, 2014). Given this structure, I hypothesized that object-object binding in scenes could be analogous to feature-binding within objects, which has been extensively studied in both VWM (e.g., Luck & Vogel, 1997; Johnson, Hollingworth, & Luck, 2008; Karlsen et al., 2010; Luria & Vogel, 2011; Brady & Tenebaum, 2013; Ecker et al., 2013) and LTM (e.g., Starns & Hicks, 2005, 2008; Meiser & Bröder,
One major question that comes out of the feature binding literature is whether features of objects are bound to each other as well as to the object to which they belong. As a result, the analogous question in scene memory is whether objects in scenes are bound to each other in addition to the scene to which they belong (see Fig. 1). In particular, it seems that feature binding is not an “either or issue”, but rather in which circumstances features bind directly to each other, or remain independent of each other. That is, the extent to which bound-feature or independent-feature representations are found depends on how attention is directed to the feature relationship at encoding (e.g., Woodman & Vogel, 2008; Fougnie et al., 2010; Troyer et al., 1999; Boywitt & Meiser, 2012). Therefore, it is plausible that direct object-object binding in scenes will occur when those objects were well attended at encoding.

Provided that direct object-object binding is possible for attended objects, another question that arises is the degree to which two objects are bound to each other. Analogous feature-binding research suggests that binding strength between features varies (e.g., Brady & Tenebaum, 2013; Orhan & Jacobs, 2014; Meiser, 2014), so it is likely that object-binding in scenes does so, too. This could happen for a number of reasons, one being that the focus of attention shifts rapidly as scene information is encoded. As a result, direct binding among attended objects may vary on a number of environmental factors that influence attentional focus, either overtly or covertly. In this study, spatial and temporal (defined as the order in which objects were attended) proximity among objects were identified as likely factors that influence binding strength, based on
research in contiguity effects. For example, in list learning it is well documented that words studied in close proximity tend to be clustered together in memory (e.g., Howard & Kahana, 2002a, 2002b). A similar phenomenon arises in map learning such that locations on a map that are studied in close proximity, or are spatially closer to each other on the map, tend to be remembered in clusters (e.g., Kosslyn, 1978; McNamara, et al., 1984). This grouping of words and map locations is a form of binding and therefore accuracy on these types of memory tasks can be used as a measure of binding strength. When this is done, there is a clear relationship in proximity such that memory for these associations get weaker as the stimuli become more distal from each other, despite focused attention to each stimulus.

In Experiment 1, I applied what is known about feature-binding in objects to hypothesize about object-binding in scenes. I tested the prediction that focused attention on certain objects in a scene (but not others) would facilitate direct object-object binding for the objects that were the targets of focused attention at encoding. Further, it was predicted that if direct object-object binding occurs in addition to object-context binding that this type of binding would be observed regardless of the utility of the context, but would be exaggerated in useful contexts. That is, the similar contexts, which were designed to cause interference and disrupt object-object binding, would not fully eliminate evidence of direct object-object binding. This prediction is in line with work that shows direct binding of features within objects, when those features are well attended at encoding (e.g., Meiser & Bröder, 2012). On the other hand, if objects were only related to each other through their independent relationships with the broader scene context, then
evidence of direct binding should have only surfaced for object pairs studied in the unique contexts, but not the similar contexts. This alternative prediction is in line with independent accounts of feature relationships (e.g., Starns & Hicks, 2008; Bays et al., 2009).

The first hypothesis that attended objects would show evidence of direct binding (via better memory for those pairs) was supported. Attended object pairs were remembered better than unattended object pairs, regardless of the context in which they were studied. This is evidence of direct object-object binding within a scene because this result occurs regardless of the type of context in which the objects were studied. If object-object binding was mediated by object-context memory, there should have been no evidence of binding for object pairs in the similar contexts. Instead, these results are supportive of other research that demonstrates that direct binding of visual information can occur provided that attention is properly focused on the to-be encoded stimuli. Additionally, these results support the analogy that objects in scenes behave like features in objects, as object relationships seem susceptible to the same manipulations as feature relationships.

The second hypothesis was not supported. It was predicted that the unique context might provide an additional boost to memory for the attended object pairs, as the unique contexts should have been able to serve as cues in ways that the similar contexts and arrays could not. It was not anticipated that there would be any effect of context on unattended object pair memory, as memory for unattended objects was predicted to be uniformly poor. In Experiment 1 context type did not influence memory for the
attended pairs, but instead influenced memory for the unattended pairs. While the participants performed below chance on the unattended test trials, performance was systematically worse for the pairs in the similar contexts and the arrays, relative to the pairs in the unique contexts. In addition to a larger proportion of hits it is possible that this effect arose from a recall-to-reject strategy, as participants correctly rejected more unattended lures from the unique contexts, than from the similar and array contexts. The unattended items, while not intentionally attended to at encoding, could have been encoded while participants searched for the attended items. As a result, the test items may have served as “good enough” cues to recall the context in which the objects appeared, allowing a rejection of mismatched pairs (and confirmation of matched pairs). If so, then this would be evidence for object-context binding that was not directly observed in the attended pairs, as the object-context relationships were used to successfully complete the associative recognition task.

Experiment 1 provided evidence of direct object-object binding in scenes dependent on the modulation of attention at encoding. In Experiment 2, I sought to determine whether binding strength differed among attended objects, based on temporal and spatial proximity. Experiment 2 tested the prediction that items that were attended in close temporal (i.e., studied one right after the other) and spatial proximity would likely be better remembered, relative to item pairs that were temporally and spatially distant. This would suggest that object pairs that were attended close in time to each other, and were also physically close to each other, would be better bound than objects that were farther away in attention time and space. Given prior research (e.g., Sherman & Lim,
1991), it was unclear whether spatial distance would produce an effect on memory independent of temporal distance. Therefore, it was hypothesized that stronger binding based on closer temporal proximity was more likely to be observed. Instead, the results revealed no significant main effects or interactions. There was an interaction between the type of context in which objects were studied and spatial distance that trended towards statistical significance. The pattern of this data revealed that close spatial proximity and context positively influenced memory for object pairs, relative to far proximity and no context (i.e., the array) conditions. This suggests the possibility that spatial proximity can influence binding between objects in memory, but is particularly useful when contextual information is present. One possibility is that spatial proximity among the objects was highlighted via the background information in the scene context. Future research is needed to examine this hypothesis. Additionally, there was no evidence that binding is based on temporal proximity at all. This lack of an effect was unexpected, but could be due to task demands such that spatial proximity served as a better cue for the experiment’s visual encoding task (e.g., Curiel & Radvansky, 1998).

Experiment 1 provides preliminary evidence that demonstrates how objects in scenes are related to each other, and the broader scene context. From Experiment 1, it is clear that objects are able to directly bind to each other when they are the focus of attention at encoding. This is analogous to work on feature binding which demonstrates the same finding. In addition to direct object-object binding, it was expected that object-context binding would be observed in the form of an interaction between attended object pairs and context type. Instead, object-context binding was actually observed in the un-
attended object pairs. These results suggest that the influence of object-context relationships was overshadowed by the relationships among the attended object pairs, but proved useful when object relationships were poorly encoded (i.e., the unattended pairs). Experiment 2 provided no significant main effects or interactions. The pattern of data suggests that attended pairs that were spatially proximal were remembered better when studied in a unique, detailed context. At first, this seems at odds with the results of Experiment 1, which showed no additional benefit to attended object pair memory as a result of being encoded within a unique context. However, Experiment 1 did not account for the spatial proximity of the different attended object pairs. It is possible that an effect of context may not arise without manipulating spatial proximity.

Overall, the current work advances what is known about scene structure in LTM through determining whether objects are bound directly to each other, or only their broader scene. But the current study has some limitations. For example, Experiment 1 demonstrates the importance of attention on direct object-object binding in scenes. However, object-context binding was only observed in the unattended pairs of Experiment 1. But, without further research this cannot be confirmed. Ideally, Experiment 1 would be replicated with better stimuli, and better control of attentional focus at encoding. While the stimuli were independently labeled by several research assistants blind to the task, it was clear from participant feedback that they felt that some of the objects were difficult to identify at test, when no context was provided. Additionally, attention could have been better controlled. The questions that were presented at encoding that were intended to serve as attention checks do not seem to have had the intended ef-
fect. Even so, when analyses included only data from participants that scored well on these questions, the pattern of results remained the same. This suggests that these questions simply were not optimal for assessing whether participants were attending to objects in the way I had intended. A better attention check would be to use a recognition task that is identical to the one that participants complete at test. That is, to have participants make a “yes/no” judgment about whether an item was in the just-studied scene. This would better map on to the type of questions that participants experience at test and would likely serve as a better way to determine which participants were paying attention to the study phase. Finally, another way to better control attentional focus at encoding would be to replicate this study in-person, in a distraction-reduced environment. While many studies in attention have been conducted on Mturk, it is possible that remote data collection was not the best platform for this particular study. I conducted a pilot study prior to the current study in person and the results of that were in line with the predictions that are borne out of the literature, performance was better, and the effects were more robust. By making these changes to Experiment 1, I think that the data would provide stronger evidence for direct object-object binding in scenes.

In addition to these changes to the design, future research on object binding could include the use of eye tracking. This would provide a more direct measure of attention. It would also allow the scene stimuli to be explored by the participant in a more naturalistic way, which might change the way in which information is bound in a memory. Eye tracking is ideal for this type of design. The eye tracker would tag each attended object based on the dwell time (i.e., amount of time spent looking at an object)
for each object. This design would ensure that participants are tested on information they have actually studied and would allow an exploration of how binding occurs without overt attention direction. This type of method could also explore what characteristics people attend to in environments. The current study used spatial and temporal distance as a way to measure binding variability. A better measure might be to use participant driven characteristics to further test the idea of binding variability.

In sum, the implications of this research influence how we think about and interpret our memories. When we remember events from our lives, we often have the sense of re-experiencing them in a coherent, integrated way. This study was motivated by the broad question of whether such coherence is reflective of the relationships that exist in memory representations of those events. The results of these experiments suggest that objects within scenes can be bound to one another, apart from any relationship each object may have with the context as a whole. Furthermore, the findings suggest that attention is a key mediator of object-object binding, much like it is for feature binding (e.g., Meiser, 2014). However, as this study is one of the first to systematically study object-object binding, there remain many more questions, including whether such binding varies in strength, and whether there are consistent interactions between levels in a memory representation (e.g., object, context). Altogether, work on these questions promises to shed light on the structure of event memory. Doing so will open new paths to improve memory and to predict what elements of experience are most likely to be accurately represented in LTM.
References


Luria, R. & Vogel, E. K. (2011). Shape and color conjunction stimuli are represented as bound objects in visual working memory. *Neuropsychologia, 49*(6), 1632-1639.


Appendix A. IRB Approval

ACTION ON EXEMPTION APPROVAL REQUEST

TO: Kacie Mennie
Psychology

FROM: Dennis Landin
Chair, Institutional Review Board

DATE: November 9, 2017

RE: IRB# E10753

TITLE: Memory for Complex Scenes


Review Date: 11/3/2017

Approved X Disapproved

Approval Date: 11/4/2017 Approval Expiration Date: 11/3/2020

Signed Consent Waived?: No

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

LSU Proposal Number (if applicable): Nancy Landin

Protocol Matches Scope of Work in Grant proposal: (if applicable)

By: Dennis Landin, Chairman

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

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

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

77
Appendix B. A’ Analyses

Given the relatively low hit rate, A’ was also calculated. Results using A’ as a measure of discriminability (and therefore, memory) should parallel the results using d’. Unlike d’, A’ cannot be negative, but instead varies from 0 to 1, with larger A’ scores indicative of better performance. Again, there was a significant main effect of attention, $F(1,87) = 86.06, p< 0.001, \eta_p^2 = 0.50$. As can be seen in Figure 12, objects pairs that were attended at encoding ($M = 0.64, SE = 0.02$) were better remembered than object pairs that were not attended at encoding ($M = 0.39, SE = 0.02$).

![Mean A'](image)

Figure 12. Main effect of attention on memory for object pairs.

The interaction between attention and context-type was, again, significant, $F(2,87) = 3.29, p< 0.05, \eta_p^2 = 0.07$. Follow up tests revealed that there were no significant differences across context types when both objects in the pair were attended at encoding, $F(2,87) = 0.43, p = 0.65$. However, there was a marginally significant difference in memory performance for pairs that were not attended at encoding across the different context types, $F(2, 87) = 2.6, p<0.08$. As can be seen in Figure 13, regardless
of the type of context object pairs were presented in, participants performed below chance when determining whether the unattended pairs were old or new. However, multiple comparisons (Bonferroni corrected) revealed a marginally significant difference in memory for unattended pairs that were presented in the unique and array contexts ($p = 0.10$). Unattended object pairs in the array contexts ($M = 0.35, SE = 0.04$) were more difficult to retrieve than in the unique contexts ($M = 0.47, SE = 0.04$).

![Figure 13. Pair type x study context type interaction.](image)
Vita

Kacie Mennie attended the University of Michigan- Dearborn for her undergraduate career. Here she became involved in numerous research projects. After graduating with her B.A., she began working with Drs. Sean Lane and Jason Hicks at Louisiana State University where she has refined her interests in attention and long-term memory. She received her M.A. from LSU in 2015. She completed her dissertation work under the advising of both Sean Lane and Jason Hicks.