What's Your Source: Evaluating the Effects of Context in Episodic Memory for Objects in Natural Scenes

Pernille Hemmer (pernille.hemmer@rutgers.edu) [1]

Kimele Persaud (kimele.persaud@rutgers.edu) [1]

Rachel Venaglia (venaglir@lafayette.edu) [2]

Joseph DeAngelis (joseph.deangelis@rutgers.edu) [1]

Department of Psychology, Rutgers University [1]
Department of Psychology, Lafayette College [2]

Abstract

It is well known that the context of a scene can have a strong effect on the identification of objects in the scene (e.g., Biederman, 1972). However, it is unclear what role global versus local context plays on episodic memory for objects. We present results from a series of experiments that evaluate the degree to which the global and local context contributes to memory performance: partial scene context, where global context was partially removed, no-spatial scene context, where the local spatial relationships among objects was distorted, and random context, where the associative relationship among objects was altered. Study time was also manipulated. We compare the findings to memory performance for objects in natural scenes (Hemmer & Steyvers, 2009; Steyvers & Hemmer, 2012). Results show that background context of a scene is important for initial scene interpretation. In addition, associative and spatial context is important for the retention of a larger number of objects in memory.

Keywords: Context; Episodic Memory; Prior Knowledge; Natural scenes; Objects in scenes.

Introduction

The context of a natural scene can be defined as the structure and permanence (i.e. stable over time) of relationships among the background and objects in an environment consistent with the real world. The context of a natural scene has important implications for a variety of cognitive tasks including: how we visually search a scene, how we categorize objects in a scene, scene perception, memory for objects in a scene, attention, etc. However, the functional components of context that influence these areas of cognition are not yet fully understood.

The context of natural scenes is thought to be both globally and locally structured (Galleguillos & Belongie, 2010; Torralba, 2003). The global context (also known as scene centered context) refers to the overall configuration of a scene. Global context is responsible for the unification of objects and the background (see Figure 1a for an example of a scene with global context preserved and partially removed) and supports quick high-level semantic interpretation of a scene (Potter, Staub, Rado, & O'Connor, 2002; Torralba, 2003). It also affords individuals the ability to make predictions about objects that prototypically accompany the scene type (Galleguillos & Belongie, 2010).

Similarly, local context (a.k.a. object centered context) refers to the relations among objects in the entire scene, or in a particular region of a scene. Local context is derived from the associative and spatial arrangements of the scene objects. These associative and spatial relationships enhance scene perception and aid object recognition (Biederman, 1972; Biederman et al., 1982; Galleguillos & Belongie, 2010; Palmer, 1975; Torralba, 2003). Take for example the ambiguous object in Figure 1b (left panel). The intrinsic properties of this object make confident object recognition quite difficult. However, when that same ambiguous object is placed next to another object (i.e., the trashcan in the next panel), object recognition occurs more readily.

The spatial relationship among objects in a scene also contributes to object recognition. Figure 1c (left panel) demonstrates an ambiguous object placed above a table, which leads one to conclude that the object is a table cloth. However, in Figure 1c (middle panel), that same object is placed below the table and now one may conclude that it is an area rug. In Figure 1c (right panel), the same object is placed above a bed and now appears to be a blanket. Taken together, the context of a scene is comprised of the global level, which unifies the objects and the background of the scene, and the local level which determines the associative and spatial relationships of objects.

Although a great deal of research suggests that global and local features of scene context facilitates object identification, object categorization, and scene perception, it remains unclear how these components of context influence long-term episodic memory for objects in scenes. In this study, we employ novel stimuli, in which we systematically remove both global and local contextual features from simulated natural scenes, to measure their influences on memory. The contribution of this study is that it bridges the influence of scene context effects on object perception, object categorization, and computer vision to long term episodic memory. Previous research of scene context effects has either employed short term or working memory in an effort to understand scene perception (Biederman, 1972; Hollingworth & Henderson, 1998) or failed to demarcate the influence of global and local components of context (Hemmer & Steyvers, 2009; Steyvers & Hemmer, 2012).

We extend the work of Hemmer and Steyvers (2009) which measured long-term episodic memory for objects in full context natural scenes (figure 1a, left panel) in an effort to investigate the influence of the functional units of natural scene context on episodic memory. We outline the methodology in their study as it serves as a basis of

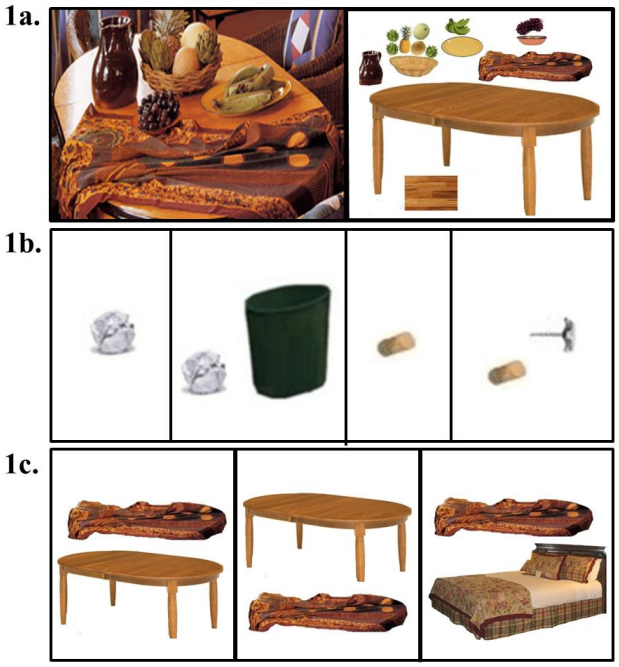

Figure 1. 1a shows a natural scene – a dining room – with the global context preserved (left) and removed (right). 1b shows the associative relationship among objects. 1c shows the spatial relationship among objects.

comparison for the contextual manipulations in the current study. Hemmer and Steyvers first measured people's prior expectation for objects in natural scenes, i.e., objects consistent with the context of that scene. Participants were asked to list all of the objects that make up each of five natural scene types (i.e., kitchen, dining, hotel, urban, and office). Interestingly, by simply guessing with their context based expectation of objects in scenes, subjects achieved a high degree of accuracy when responses were scored against actual images. Next, Hemmer and Steyvers tested free recall for the same natural scenes as a function of study time. They found that for initial responses, accuracy was highest for short study times, however, at later responses, accuracy was best for longer study times. Investigating the influence of global and local context may provide insight for the accuracy shifts across the study time conditions.

Experiments

In three experiments, we systematically evaluate the effect of contextual relations on episodic memory. In addition, we manipulate study time to determine if decreasing available context can be compensated for by additional study time. In the first study, we evaluate the effect of global context by removing background information. In the second study, we evaluate the effect of simulated spatial context by disrupting the natural spatial relationship among objects. In the third study, we evaluate the effect of associative context by presenting study items, drawn from different natural contexts, together as a scene.

We evaluate performance in all three experiments against the prior expectation experiment in the original Hemmer and Steyvers (2009) study. That is, we evaluate memory performance relative to the prior knowledge condition where performance was based only on contextual knowledge and expectations, and not episodic memory. We predict that accuracy will decrease as a function of the incremental exclusion of global and local context.

General Methods

Materials

To create the stimuli for the three experiments, we used the original images along with the ground truth assessment for what objects really occurred in the scenes from Hemmer and Steyvers (2009). Every object named in the ground truth was then cropped from the original image using Paint and Photoshop. If an object could not be clearly cropped out of an image, Google Images was used to find an object that closely matched the original scene. A ratings panel of three students was employed to measure agreement of the Google items to the original items (across all experimental conditions raters agreed 97%). Individual objects were then placed on a white background. The ordering scheme for placing the object onto the background differed for each of the experimental conditions.

Extensive measures were taken to ensure that the context manipulated scenes matched the original scene as closely as possible, which included: making the most salient objects in the original scene the most salient object in the context-disrupted scene, matching up the sizes of the objects, making the objects as clear as possible, matching up the colors and angles of the objects as closely as possible. The size and saliency of the objects was especially important when creating the context manipulated scenes because people are known to have a "normative viewing size" preference for a given object (Konkle & Oliva, 2007). Figure 2 shows an example of a partial scene context image (global context partially removed) used in Experiment 1, compared to a full natural scene.

Response Normalization

Responses for all experiments were corrected for spelling, plurals, capitalization, and qualifiers (e.g., numbers, color, size and location). For example, "chair" and "chairs" were mapped to the single entry "chair", and "silver car" was

Figure 2. Left: full natural dining room scene. Right: partial-context dining room scene, in which the background was removed.

mapped to "car". All short form responses were corrected to the full word (e.g., mayo was mapped to mayonnaise, and fridge was mapped to refrigerator). All responses that could not be interpreted were removed. The correction rules were automated and applied to all datasets uniformly.

Ground Truth

To measure performance in all experiments reported here, we checked whether a recalled object was part of any of the responses given by participants in the original ground truth assessment[1] of Hemmer and Steyvers (2009). In the original study, the ground truth was measured by asking subjects viewing a single image at a time from the image set to "report as many objects as you can see". If a response given in our memory studies matched the ground truth, it was scored as a correct response. If it did not match, we manually checked whether the recalled object could still be considered as a description of an object that was part of the image. This ground truth was then added to the original list of ground truth objects. Only if the response still did not match was it scored as incorrect.

Experiment 1: Partial scene context

We sought to investigate the effect of global context – in the form of scene background- on episodic memory. That is, to what extent does the natural setting of a room (e.g., the walls and ceiling) contribute to successful memory? We predicted that this absence of global context would disrupt gist extraction and result in a decrement in performance. We expected that additional study time would improve memory performance, and sought to evaluate both the effect of removing the background context, and the amount of time needed to return to equivalent recall performance levels of a non-disrupted scene. We will refer to this as the partial-context condition.

Participants

Fifty-three undergraduate students at Rutgers University participated in exchange for either course credit or monetary compensation of $10.

Materials and Procedure

To create the partial-context images, objects were placed onto the white background in the same spatial organization as the objects in the original image. The 10 images were used to form two sets of five images, one from each scene type (kitchen, dining room, office, hotel room and urban scene). We followed the exact experimental procedure of Hemmer and Steyvers, and employed a recall paradigm in which images were presented at the center of the computer screen for either 2, 10 or 15 seconds. A simple 'find 5 mistakes' picture distracter task was inserted between study and test trials. At test, participants were asked to list all the objects they could recall from the image presented in the preceding study trial. Participants were given clear verbal instructions to ensure that they understood the task.

On study trials, study times were randomly assigned as either 2, 10 or 15 seconds following a Latin square design. On test trials, participants were required to type responses or wait 60 seconds before they could move to the next study trial. Each participant only saw 5 images, one from each scene type, to avoid carryover effects where the memory from one scene type affects recall of another image of the same type. The 5 images were presented in random order. At the end of each of the five test sessions, participants received feedback on the number of correct responses, and how many more objects they could have recalled.

Results

Performance was measured in terms of mean accuracy as a function of the output position (i.e., the order in which responses were given) and study time (See Figure 3, left panel). Figure 3, also includes the results from the prior knowledge condition in the Hemmer and Steyvers study, as a baseline for comparison. Because subjects were allowed to determine the number of responses they wanted to provide, the number of responses for each output position varied (see Figure 5 for the average number of outputs by study time and condition). Therefore, we restricted the analysis to the first five output positions. A 5 (output position) x 3 (study time) repeated measures with-in subject ANOVA was conducted to evaluate the effect of study time and output position on recall accuracy. There was as a significant main effect of output position (F [4, 272] = 17.97, p<.001), such that greater accuracy was achieved in initial output positions relative to later output positions. There was also a significant main effect of study time (F [2, 272] = 8.92, p<.001), such that greater accuracy was achieved for longer study times compared with shorter times. Lastly, there was a significant interaction between output position and study time (F [8, 272] = 4.92, p<.001). Overall, mean accuracy decreased as a function of output position. However, the removal of the background context negatively affected the 2 second condition compared to the 10 and 15 second conditions. This might be due to disturbed global level gist extraction with the removal of the background.

Experiment 2: No spatial context

While the partial removal of global context only appears to have a negative impact on short study time conditions, other aspects of scene context have been shown to disrupt memory performance for natural scenes (Biederman, 1972). In the next experiment, we tested the influence of spatial context on episodic memory for scenes by placing the objects in Experiment 1 in random order on a white background. We predicted that this absence of spatial context would disrupt the ability to use local spatial context, and result in a decrement in recall performance across study times. We refer to this as the no-spatial context condition.

Participants

Fifty Rutgers University undergraduates participated in exchange for either monetary compensation of $10 or course credit, and were not involved in Experiment 1.

[1] The ground truth of the occurrence of an object in the given images is stationary and therefore we did not replicate this portion of the study. See Hemmer and Steyvers, 2009 for further detail.

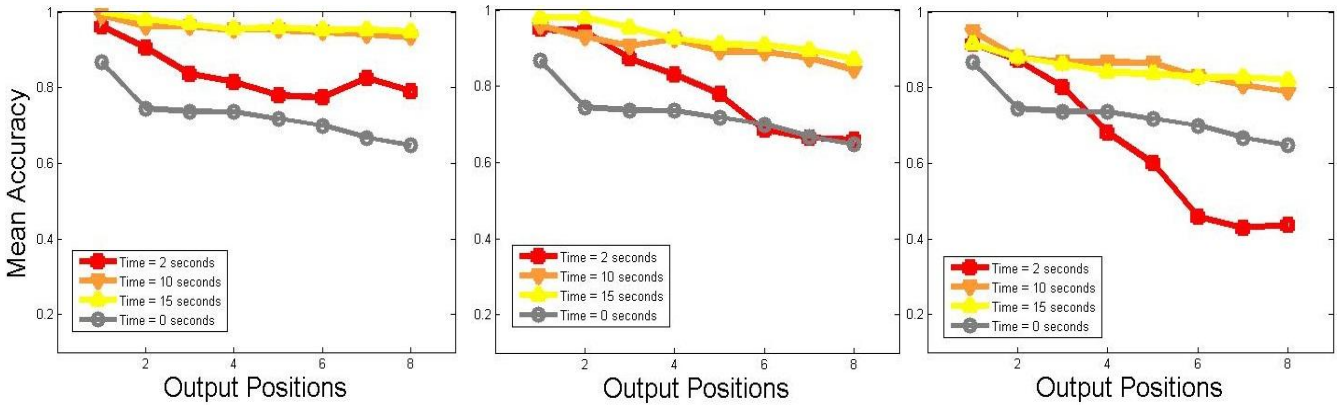

Figure 3. Mean accuracy as a function of output position and study time. Each line gives a different study time condition. The gray line gives performance from the Prior Knowledge condition of Hemmer and Steyvers (2009). Left panel: partial-context condition. Middle panel: no-spatial context condition. Right panel: random context condition.

Materials and Procedure

The materials were identical to those used in Experiment 1, except the spatial relationship among objects was not retained. Instead objects within each scene were placed onto the white background in random order. Figure 4, left panel, shows a sample image from the no spatial context condition. The procedure was identical to Experiment 1.

Results

As in Experiment 1, performance was measured in terms of mean accuracy as a function of output position. Figure 3, middle panel, shows accuracy across output positions and study time, as well as the prior knowledge condition for comparison. As in Experiment 1, we restricted the analysis to the first five output positions. A 5 (output position) X 3 (study time) repeated measures with-in subject ANOVA was conducted to evaluate the effect of study time and output position on recall accuracy. There was as a significant main effect of output position (F [4, 200] = 17.59, p<.001), resulting in greater accuracy for initial output positions relative to later output positions. There was also a significant main effect of study time (F [2, 200] = 3.36, p<.05), such that greater accuracy was achieved for longer study times compared with shorter times. There was no significant interaction. Overall, accuracy decreased as a function of output position. In contrast to Experiment 1, the removal of spatial context negatively affected accuracy in all 3 study time conditions. Performance in the 2 second condition was no better than guessing with prior knowledge after 5 output positions.

Experiment 3: Random scene context

The preceding two studies revealed a continuous decline in memory performance with the progressive removal of context information in a natural scene. In the next experiment, we tested memory for random objects. The inclusion of a random context condition served to quantify pure episodic memory and allowed additional comparisons of the influence of prior knowledge for naturalistic stimuli. We predicted that this absence of natural context would result in a further decrement in performance across study times. We refer to this as the random context condition.

Participants

Forty-eight undergraduate students from the Rutgers University participated in exchange for either monetary compensation of $10 or course credit. These participants were not involved in Experiment 1 or 2.

Materials and Procedure

The materials were identical to those used in Experiment 1 and 2, except the scene context among objects was not retained. Instead, objects within each study set were drawn at random from across the 5 scene types, and placed in random order on the white background. In this way, the stimulus no longer retained the global or local context of a natural scene.

Objects were matched for size from the stimuli in Experiments 1 and 2, such that small and large objects occurred in all scenes, and no one object was allowed to repeat across the 5 random scenes in each set. Again, a three person-rating panel was used to determine the consistency of the overall quality of the 'scene' relative to the previous experimental stimuli. Figure 4, right panel, shows a sample image from the random context condition. The procedure was identical to Experiment 1 and 2.

Results

As in Experiment 1 and 2, performance was measured in terms of mean accuracy as a function of output position. Figure 3, right panel, shows accuracy across output position and study time, as well as the results from the prior knowledge condition as a baseline for comparison. As in Experiment 1 and 2, we restricted the analysis to the first five output positions. A 5 (output position) X 3 (study time) repeated measures with-in subject ANOVA was conducted to assess the effect of study time and output position on recall accuracy. There was a significant main effect of output position (F [4, 144] = 10.61 p<.001), in the form of greater accuracy for initial output positions. There was also a significant main effect of study time (F [2, 144] = 8.38, p<.001), where greater accuracy was achieved for longer study times. Finally, there was a significant interaction between output position and study time (F [8, 144] = 5.59, p<.001). As a whole, accuracy decreased as a function of

Figure 4. Left panel: No-spatial context dining room scene in which the spatial relationship between the objects was disrupted. Right panel: Random context scene, in which objects were randomly selected from the 5 images in each scene set to create the random study list of images.

output position. As in Experiment 2, the lack of coherent context negatively affected accuracy in all three study time conditions, but more so for the 2 second condition.

Comparison of Results across All Experiments

The results from the three experiments appear to show a proportional decline of memory accuracy with the successive removal of scene context. To evaluate this effect, three repeated measures ANOVAs were conducted to compare performance for each study time (i.e., 2, 10, and 15 seconds) across the 3 contextual manipulated conditions. For the 2 and 15 second study times, the 5 (output position) X 3 (context conditions) repeated measures ANOVAs revealed a significant main effect of context conditions (2 sec.: F [2, 144] = 7.99, p<.0013; 15 sec.: F [2, 144] = 10.10 p<.001)]. However, the main effect of context condition for 10 second study time was not significant (F [2, 144] = 0.51, p=0.60)]. There was also a main effect of output position [2 seconds: (F [4, 144] = 38.64, p<.001); 10 seconds: (F [4, 144] = 8.84, p<.001); 15 seconds: (F [4, 144] = 5.06 p<.001)]. These findings suggest that the removal of context negatively affects memory performance, especially at short study times.

In addition, Figure 5 shows the number of responses as a function of context condition and study time. A 3 (study time) X 3 (context condition) repeated measures ANOVA

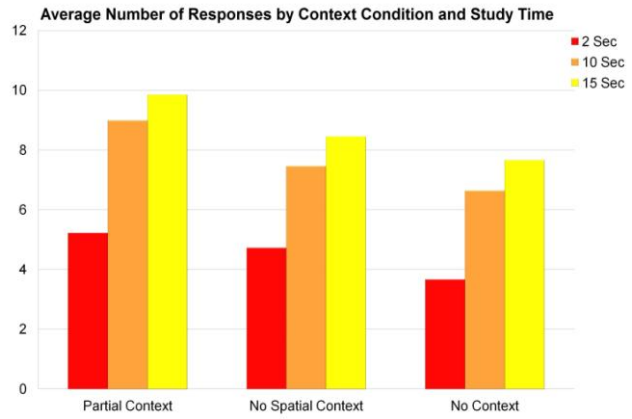

Figure 5. Average number of responses by context condition and study time

was conducted to evaluate the effect of study time and context condition on number of responses. There was a significant main effect of study time (F [2, 294] = 210.72, p<.001), with more responses for longer study times. There was also a significant main effect of context condition (F [2, 294] = 45.36, p<.001), with more responses for more context. Lastly, there was a significant interaction between study time and context condition (F [4, 294] = 6.69, p<.001), such that there were more responses given in the 10 and 15 second conditions compared to the 2 second condition as a function of the context manipulations

We also evaluated the time needed to compensate for the decrement in performance with decreasing available context. Figure 6 shows mean accuracy for each output position as a function of study time. We compared performance to a full context natural scene (Hemmer & Steyvers, 2009), using the result from the 5th output position, where accuracy in the 2 and 10 second conditions were equivalent (black dashed line). With the partial global context removed (left panel), performance for both the 10 and 15 second conditions remained above the full context performance by the 5th output. While it might appear somewhat counterintuitive that removing global context information helped performance, this might be due to the additional 'clutter' that global context adds. In contextually manipulated scenes, there were a limited number of objects available, and this constraint appeared to help performance. For the no-spatial context condition (middle panel), performance for both 2 and 10 second study time fell below full context performance on the 5th output, while the 15 second condition was slightly above. Lastly, in the random context condition (right panel), only the first output position was above full context performance for the 5th output. Even the 15 second study time was not enough to compensate for the loss of both global and local information.

The types of intrusions participants made further elucidated the distinction in contribution of global and local context to memory. Under the influence of global context, participants incorrectly recalled objects that were consistent with the overall scene and prototypically accompany that scene type. For example, when participants studied office scenes, they falsely recalled calculator because this object is highly representative of objects that are generally found in natural office settings. Similarly, while under the influence of local context, participants inaccurately recalled studying objects that are typically found in close proximity of some objects that were present in the scene. For example, when participants studied urban scenes and saw a sky, they falsely recalled seeing clouds because clouds are often found in the sky in natural urban settings.

Discussion

In this paper, we assessed the relative contribution of global and local components of context to recall for objects in natural scenes. First, we partially removed the global scene context (i.e., removal of the background, walls, and ceiling) and found a decrement in memory performance for

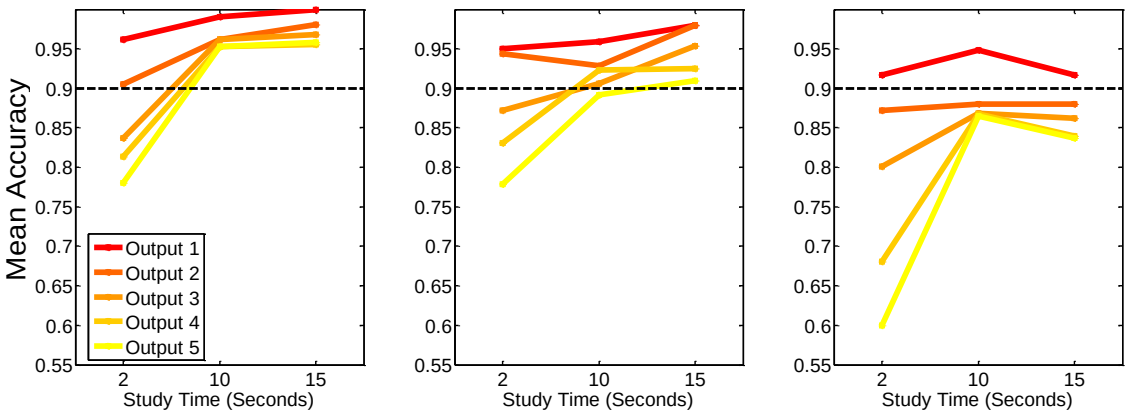

Figure 6. Mean accuracy as a function of study time. Each line gives a different output position. The dashed black line gives performance from a full natural scene context – 5th output position (Hemmer and Steyvers, 2009). Left panel: partial-context condition. Middle panel: no-spatial context condition. Right panel: random context condition.

shorter study time conditions. Previous studies have found that the global context supports quick interpretation of a natural scene, which is important for performance at shorter study times, but is compensated for by the available local context at longer study times. In the next two experiments, we systematically removed local context by removing the spatial relationships of objects (i.e. randomize the locations of objects in the scene) and the associative relationships of objects (i.e. randomize objects from various scene types). While the removal of the background of a natural image initially impedes memory for short study times, the removal of spatial and associative context impinges on both short and long study times. The results of the analysis for the number of responses further illustrate the benefit of having global and local context where, the mean number of responses for full natural scene context > partial context (global context partially removed) > no-spatial context > random context (associative context removed). The same trajectory applies to study time.

These finding have implications for our understanding of the effect of global and local context on long-term episodic memory. The removal of global context affects quick scene interpretation, and for memory this impacts performance for shorter study times, but is compensated for, at longer study times, by the available local context. However, the removal of spatial and associative information in local context affects memory performance for both short and long study times. Taken together, these studies show that global context is important for scene interpretation, which initially helps memory, but local context is important for sustained memory performance.

Acknowledgments

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant Number NSF DGE 0937373, National Science Foundation I.G.E.R.T. under Grant Number NSF DGE 0549115, National Science Foundation REU under Grant Number IIS-1062735, and Rutgers University Aresty Summer Science Program.

References

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77-80.

Biederman, I., Mezzanotte, R.J., & Rabinowitz, J.C. (1982). Scene Perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143-177.

Galleguillos, C., & Belongie, S. (2010). Context based object categorization: A critical survey. *Computer Vision and Image Understanding,* 114(6), 712-722.

Hemmer, P., & Steyvers, M. (2009). Integrating episodic and semantic information in memory for natural scenes. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Hollingworth A., & Henderson, J. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General 127*(4), 398-415

Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in object perception and memory. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society, Austin, TX: Cognitive Science Society.*

Palmer, S. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519 526.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. Journal of Experimental Psychology: *Human Perception and Performance, 28,* 1163-1175.

Steyvers, M. & Hemmer, P. (2012). Reconstruction from memory in naturalistic environments. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation*, (pp.126-144). Elsevier Publishing.

Torralba, A. (2003). Contextual priming for object detection. *International Journal of Computer Vision, 53,* 169-191.