Steyvers, M. & Hemmer, P. (2012). Reconstruction from Memory in Naturalistic Environments. In Brian H. Ross (Ed.) The Psychology of Learning and Motivation, Vol 56. Elsevier Publishing, pp. 126-144.

CHAPTER FOUR

Reconstruction from Memory in Naturalistic Environments

Mark Steyvers *and* Pernille Hemmer

Contents

Abstract

Many aspects of our experiences do not have to be explicitly remembered, but can be inferred based on our knowledge of the regularities in our environment. In this chapter, we investigate the interaction between episodic memory and prior knowledge in naturalistic environments. In contrast to previous studies that suggest a detrimental effect of prior knowledge, we show that when using stimuli that are statistically representative of our environment, prior knowledge of the regularities of our environment can lead to very different outcomes. For example, simple "guessing" using prior knowledge alone—without using episodic memory—leads to relatively high accuracy. In addition, we find relatively few intrusion errors in studies involving natural scenes. We argue that it is important to use ecologically valid stimuli in memory studies, because the findings of memory studies using statistically unrepresentative stimulus material are unlikely to give insights about the operation of human memory in more natural settings.

Psychology of Learning and Motivation, Volume 56

ISSN 0079-7421, DOI 10.1016/B978-0-12-394393-4.00004-2

1. Introduction

General knowledge and prior expectations about events are rich sources of information that are known to exert strong influences on memorial processes. Bartlett's (1932) seminal research on reconstructive memory demonstrated how cultural and social norms, as well as cognitive expectations, influence our recall of past events. Recalling events, such as a recent visit to a colleague's office might be partially based on episodic memory for the event, but also on general knowledge and expectations about the kind of objects you find in an office. You might infer that your colleague's office had books, not because you have detailed explicit memories of this object, but because offices typically have books. In this way, many aspects of our experiences do not have to be explicitly remembered, but can be inferred based on our knowledge of the regularities of our environment.

While the influences of prior knowledge have received much attention, research has focused primarily on the fallibility of memory as a result of prior knowledge. This is especially the case for research paradigms, such as studies that are designed to elicit false memory in laboratory settings. In a typical false memory study, the strongest associated or expected item is withheld from the material to be remembered. For example, in the verbal domain, creating strongly associated wordlists, and then withholding the strongest associate at study results in intrusions of highly related target words in free recall (Roediger & McDermott, 1995). Similarly, expectations about objects in scenes can lead to recall of objects that were not present in the scene. When removing an expected item, for example, books from an office, people recall seeing the expected item (Brewer & Treyens, 1981). One interpretation of these false memory studies is that prior knowledge mostly has harmful effects on the accuracy of episodic memories—we would be better off if prior knowledge would not contribute at all.

However, rather than taking this as evidence of the shortcomings of memory, the incorrect responses elicited in false memory studies might also provide some insight into the functions of memory. Recalling books in an office might be the result of an attempt by the memory system to reconstruct an event based on the knowledge that in most offices we encounter there *are* books. Falsely recalling books in the Brewer and Treyens office study is an error because the environment has been manipulated for the purpose of the experiment and is not representative of the naturalistic environment we usually encounter in offices. Brunswik (1955) argued that behavior observed in a constrained environment can only speak to behavior in that environment, and not to the general behavior of an organism in an unconstrained environment. In other words, false memory studies provide compelling evidence of memory functions when expected objects are absent, but they do not tell the complete story of the functions of memory in naturalistic—unconstrained—environments.

Adopting an approach where the stimuli are sampled from a natural environment, which is known to the subject, has led to rather different findings than those of the false memory literature. Huttenlocher and colleagues (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Prohaska, 1992; Huttenlocher, Hedges, & Vevea, 2000), in a series of studies on categorical perception, showed that having prior knowledge of the stimulus distribution improved average recall. Their studies suggested that people can use the overall knowledge of the stimulus distribution to fill in noisy and incomplete memories for events. While the stimulus material used by Huttenlocher and colleagues was in no way naturalistic (it consisted largely of dots in circles), these studies demonstrated that having knowledge of the underlying environmental regularities can serve as an aid to episodic memory and lead to improvements, rather than decrements, in memory performance.

Recent studies adopting more naturalistic stimuli have demonstrated effects similar to those of Huttenlocher et al. (e.g., Konkle & Oliva, 2007). For example, prior knowledge for naturalistic stimuli can serve as an aid to older adults (Castel, 2005). When asked to recall prices for common grocery items older adults, generally shown to have impaired memory performance relative to younger adults, performed as well as younger adults. This is thought to be a result of the meaningful naturalistic information embedded in the stimulus (i.e., grocery prices). In amnesic patients with compromised semantic systems, however, no such improvement is seen (Kan, Alexander, & Verfaellie, 2008). While healthy controls showed enhanced performance on recall for prices of grocery and household items amnesic patients did not. This suggests that semantic memory plays a key role in memory function, and that accessing and integrating prior knowledge can lead to improvements in memory performance. For example, recall can be quite accurate in situations where participants have pre-experimental prior knowledge and the stimuli follow a natural distribution, compared to situations where participants have to remember abstract shapes for which no prior knowledge has been established (Hemmer & Steyvers, 2009a, 2009b).

As evidenced by the radically different findings of the false memory studies and the prior knowledge studies, the choice of stimulus plays an important role in the findings elicited in a laboratory context. Traditionally, memory studies have focused on stimuli designed with no structure (e.g., random word lists or random sets of faces) in order to attain maximum experimental control. Ebbinghaus (1885) famously used nonsense syllables to control for the prior associations underlying memory processes. Such experiments are deliberately designed to study the memory processes independent of the stimuli. Stimuli can also be designed to be generally representative of the environment and then statistically manipulated in order to leave out items (e.g., words or objects) that are to be expected in the given environment. The Deese, Roediger, and McDermott (DRM) paradigm is a good example of this (Deese, 1959; Roediger & McDermott, 1995). In those types of designs prior knowledge is placed in opposition to episodic memory in order to study the resulting errors. While these types of designs have had a large impact in memory research, the relationship between the natural environment and memory seems to have been lost in the quest to maximize experimental control.

In this chapter, we will be presenting a point of view arguing in favor of ecological validity[1] in memory studies. We will present findings from studies using ecologically valid stimuli that are statistically representative samples from the natural environment. We will argue that it is important to incorporate memory stimuli in laboratory studies that are designed to be statistically representative samples from the natural environment, as opposed to manipulated for the purpose of experimental control. In such studies, prior knowledge about the environment and expectations about events can play a more natural role in the reconstruction of events from memory.

In sum, we will argue that memory researchers should be careful in drawing generalizations from laboratory studies involving stimuli that are unrepresentative of the natural environment. Figure 1 illustrates our general point. It shows the particular office scene used in the Brewer and Treyens study (that does not have books) and also a set of unaltered office scenes. In this chapter, we will argue that studies involving altered stimulus material—such as the majority of studies involving false memory—generalize to only a small subset of scenarios. In addition, we assert that studies involving more ecologically valid material will not only generalize to a larger and more diverse set of scenarios, the conclusions drawn from such studies might be different. In order to support these arguments, we will review some of our previous research and highlight some of our empirical findings of a study involving naturalistic scenes.

[1] The term ecological validity has changed in meaning after Brunswik's original contributions (Hammond, 1998). Here, we mean that the stimulus is unaltered, and presents a representative sample of the natural environment. This more closely resembles Brunswik's concept of representative design.

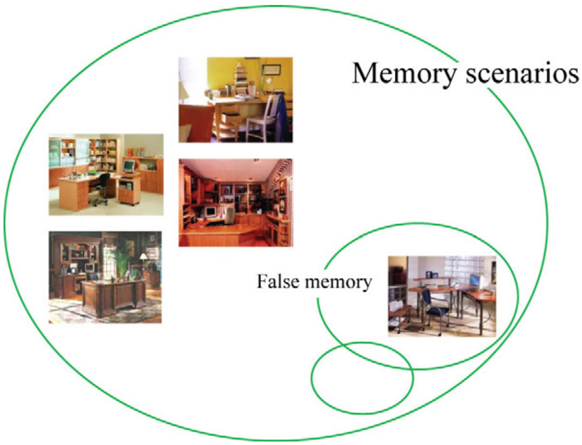

Figure 1 False memory paradigms only represent a subset of possible scenarios in which one can study memory. In many other memory scenarios the stimuli to be remembered are ecologically valid and recall performance is dramatically different than in manipulated environments. In the false memory scenario, the illustration shows an office that does not contain books. All other illustrations of offices do contain books and represent more natural samples from the environment. (For color version of this figure, the reader is referred to the web version of this book.)

2. Assessing the Influence of Prior Knowledge in Recall for Natural Scenes

In much of our previous research, we have examined the influence of prior knowledge on episodic memory using naturalistic stimuli that are representative of the environment (Hemmer & Steyvers, 2009a, 2009b, 2009c; Hemmer, Steyvers, & Miller, 2010). The results from these studies have provided converging evidence of the beneficial influences of prior knowledge on reconstructive memory. Several interesting and noteworthy findings result from these studies. First, prior knowledge can be utilized to "clean up" noisy episodic representations, thereby leading to an overall increase in accuracy in reconstruction from memory. This is true even when participants do not remember studying the objects. Second, we found that prior knowledge had effects at multiple levels of abstraction, and we proposed that these influences are hierarchically structured. For instance, recall for objects with limited categorical information (artificial shapes) was biased towards the mean of the overall distribution of artificial shapes, whereas recall for objects with clear categorical information (fruits and vegetables) was biased towards distributions associated with specific objects. Similarly, prior knowledge for height might exist not only for the general height of people, but also at a more fine-grained level based on gender (females on average are shorter than males). Thus, prior knowledge at a more fine-grained level might contribute to further improvements in average recall over general level knowledge. This interaction between episodic memory and prior knowledge is dependent on familiarity. If a participant studies an object with which they are familiar, for example, a chayote (a type of gourd), then they can use their knowledge about the common size of this object to aid their reconstruction and correct an otherwise noisy memory trace at test. Another participant that studies the same chayote, who does not know this object, might be able to recognize it as a vegetable and can use his general knowledge at the category level to guide reconstruction.

In this chapter, we will explore the consequences of using naturalistic stimuli on the study of memory and illustrate our arguments with data from one of our previously published studies (Hemmer & Steyvers, 2009c). In this study, hereafter referred to as "the scene study," we investigated memory for objects in naturalistic scenes, such as kitchens and offices (for full details see Hemmer & Steyvers, 2009c). The goal of the study was to characterize the interactions between episodic and semantic components in recall for objects in occurring in naturalistic scenes. In the remainder of this chapter, we will show how to assess prior knowledge experimentally and how to evaluate the potential benefits of prior knowledge in reconstructive memory. We will also introduce new analyses where we reevaluate standard memory findings related to object consistency, novelty, and false recall. We will assess these findings in the broader context of unconstrained environments in order to elucidate the full story of memory function.

3. Assessing Prior Expectations

An important first step in characterizing the interactions between episodic and semantic components is to quantify prior expectations. When training participants on the underlying environmental distributions of the stimuli (e.g., Huttenlocher et al., 1991) prior knowledge is explicitly designed and controlled by the experimenter. When using naturalistic stimuli for which participants have pre-experimental prior knowledge however, prior knowledge is under the control of, and dependent on, the individual participant.

In our previous studies (Hemmer & Steyvers, 2009a,b), we developed a novel method for eliciting people's prior expectations experimentally. The procedure is very simple—participants are asked to provide their perceptual judgment of some feature of the stimulus, and we take this as a measure of the prior expectation for the stimulus. People have been shown to be quite accurate when asked to make

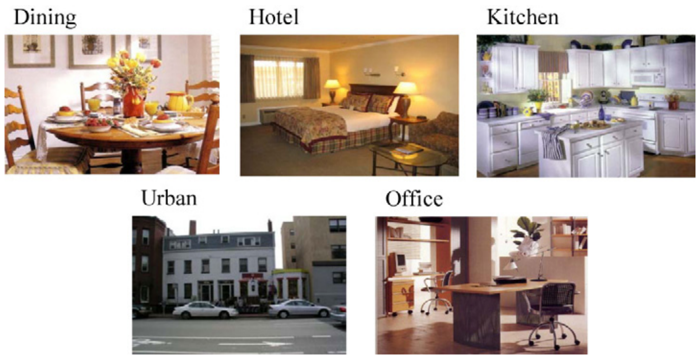

Figure 2 **Sample images from each of five scenes types used in our scene memory experiments. (For color version of this figure, the reader is referred to the web version of this book.)**

perceptual judgments based on knowledge of natural distributions, for example, estimating height based on accessible gender information (Kato & Higashiyama, 1998).

In 'the scene study" that we will be discussing in this chapter, we sampled 24 high-resolution color images from the LabelMe database (Russel & Torralba, 2008). The 25th image was the original image used by Brewer and Treyens (1981). There were five images in each of five scenes types (kitchen, office, dining room, hotel room, urban scene). See Figure 2 for sample images. To assess prior expectations participants were asked to name the objects they would expect to find in a given scene type (e.g., a kitchen). No stimulus image was presented to the participants in this verbal cue condition; the task was simply for the participant to name objects that they would expect to see in a given scene based on prior experiences. We take the resulting frequency distributions over participants as a measure of people's prior expectations for the occurrence of object in a scene. A separate group of participants was asked to provide perceptual judgments of the actual stimulus images (i.e., name all the objects in a given image). This perceptual condition is needed because some objects might objectively be present in a scene but be imperceptible to people. Therefore, we treat the resulting frequency counts from this condition as measures of the ground truth for the objects present in the scene images.

Figure 3 shows an example of the top ten responses for the verbal cue "Urban scene." It also shows the top ten responses (in the perceptual condition) for a specific image of an urban scene. Responses reflect the intuitive notion that objects that are central and salient in an image have high response frequencies. For example in the Urban Scene image, the most central and salient object is a car that was named by 20 out of 22

Figure 3 Illustration of response frequencies elicited in our norming study. Participants list objects that they see in a particular image (perception condition) or that come to mind when probed by a type of scene (verbal cue condition). The results from the perception condition can be used to assess the ground truth about objects in a scene whereas the verbal cue condition can give estimates of the prior expectations about objects in scenes.(For color version of this figure, the reader is referred to the web version of this book.)

subjects. For the prior knowledge norms (where no image is presented) responses were more generalized, and certain scene types appear to be associated with very iconic objects, for example, a computer in an office (20 out of 22 subjects), a television in a hotel room (22 out of 22 subjects), and a table in a dining room (19 out of 22 subjects). For both the Urban Scene image and the Urban Scene cue, the top three responses were the same: car, building, and people. There are however idiosyncratic differences. The specific image is that of a clothing store, and thus 18 people responded "clothing." Clothing, however, is not named for the scene cue. The scene cue on the other hand generated the response "tree," which is not present in the specific image and therefore was not given as a response. Interestingly, three participants generated the response "book" while viewing the Brewer and Treyens office image. Although we know that there were no books in the office, items in the image were perceptually similar to books. Overall the results indicate that people have strong prior expectations and that assessing these expectations experimentally provides a good description of people's prior knowledge that are in line with the occurrence in the natural environment.

4. Assessing Episodic and Prior Knowledge Components in Recall

To assess the interactions between prior knowledge and episodic memory for naturalistic scenes, the memory phase of the scene experiment used the same scenes as in the prior knowledge study. We used 10 of the 25 images from the perception phase (two from each scene type) to form two sets of five images. The images were chosen based on generating the highest number of responses in the perception study. We employed a continuous recall paradigm with the sequence of study and test trials randomly interleaved.

The scenes were shown at study for either a short (2 s) or a long (10 s) duration. This manipulation was intended to change the degree to which participants relied on prior knowledge in episodic memory retrieval. A participant who has the opportunity to extensively study a scene might simply use her episodic memory trace in reconstructing the event and not rely on prior knowledge. This should be evident by the recall of objects with no prior scene expectation, that is, objects that are not consistent with a scene, which can only be recalled episodically. On the other hand, in the condition with relatively short presentation times, participants might simply not have the opportunity to store all detail in episodic memory. In this case, we expect that participants rely more on prior knowledge to fill in the "holes" in their episodic memory—these are objects that for some reason did not get encoded but their presence in the original scene might be inferred based on general knowledge. In this memory phase, we can evaluate the influence of prior knowledge on recall for natural scenes. We next discuss three kinds of analyses of the experimental results: (1) analysis of errors such as intrusions, (2) analysis of recall for scene consistent and inconsistent objects, and (3) the influence of prior knowledge in the absence of episodic memory.

5. Analysis of Errors

In studies involving intentionally manipulated stimuli, it is clear what errors are of interest to the researcher. The main focus in those studies is on intrusions involving items that are removed from the scene (or words that are the associative of a list of words but are then removed from the word list). However, in our case, since we did not manipulate the scenes, we were interested in the types and frequencies of errors observed in our memory study.

We evaluated errors (i.e., naming an object that is not present in a given image) as a function of the prior probability of the object in a given scene. This prior probability was derived from the responses in the prior knowledge study. The overall error rate for high prior probability objects was 9%. The error rate for low prior probability objects was 18%. Not only are these error rates low, but they are counter to the finding of Brewer and Treyens who found higher error rates for high prior probability items. When the stimuli are unmanipulated and representative of the environment however, high prior probability objects are a priori likely to be present in the scene, leading to a lower error rate.

It should be pointed out that in our free recall paradigm we used a method of solicitation where participants were free to terminate responses at will. This resulted in an average output for the 2 s condition of 7.75 items whereas the average output for the 10 s condition was 10.05 items. The average cutoff threshold was 90% accuracy for both conditions. This suggests that participants monitor their response accuracy and that the average response threshold is about one in ten. This finding is consistent with the findings of Koriat and Goldsmith (1996) that participants monitor and control for accuracy in free recall.

We also evaluated the conditional error rates, that is, the recall probability of an object given that the object is absent from the scene. These are similar in nature to the false memories of the Brewer and Treyens study, such as recalling books that were not there. By virtue of the stimuli being representative of the true environment and using a free recall task there were not many such items. One clear example was the conditional error rate for "table cloth" in the dining scenes. Table cloth had a high prior probability for the dining scenes as assessed in the prior knowledge norming experiment (i.e., where participants list objects they expect to see in a dining scene). However, one of the dining images did not have a table cloth present (see Figure 4). This resulted in a conditional error rate of 19%. While this is 10% higher than the overall error rate for high-probability items, it is still not as great as that of the books in the Brewer and Treyens study, with a (false) recall probability of 30%. The striking difference between the findings of Brewer and Treyens (and the false memory literature in general) and our findings speaks of the importance of evaluating memory as a function of the natural environment. While the standard assumption has been that prior knowledge and expectations lead to intrusions in recall, we have shown that when the to-be-remembered scenes are representative of the environment, we can expect much reduced error rates. This finding could only have been brought about by using naturalistic stimuli. Errors such as falsely recalling books are very likely in a situation where we have an office without books but this situation has a low base rate—because we encounter very few offices without books, this type of error is relatively rare. In our experiment, we found very few intrusions overall mainly because objects that are consistent with a scene are a priori more likely to be present in that scene.

Figure 4 Example of a dining scene where a high-probability object such as a table cloth is missing. The intrusion rate for this object is relatively high in this particular scene. (For color version of this figure, the reader is referred to the web version of this book.)

6. The Effect of Object Consistency

Two conflicting findings regarding the effect of object consistency have emerged in the literature. Brewer and Treyens (1981) found that objects that are consistent with many scenes are better remembered. They showed that saliency was positively correlated with recall and recognition for present objects, and they showed a strong positive correlation between schema expectation and recall. For example, a high prior expectation item for a graduate student office at the time of the study—a typewriter—was recalled by 90% of participants, whereas a low prior expectation item—a skull—was only recalled by 50% of participants. From this they concluded that participants used a "room schema" to retrieve objects in the recall task. They did however, also caution that present objects with high prior expectation for a given scene might be recalled based on schematic information (i.e., prior knowledge) rather than episodic information, and that they could not distinguish between the two. Pezdek, Whetstone, Reynolds, Askari, and Dougherty (1989) in contrast argued against this conclusion and showed that novel objects (not consistent with the scene) are better remembered; also known as a novelty or von Restorff (1933) effect.[2]

[2] Somewhat confusingly, Pezdek et al. referred to this as a "consistency" effect.

Their experimental stimuli were designed such that each setting (a graduate student office or a preschool classroom) contained 16 items of which half were consistent with the setting and half were inconsistent. The items were viewed either in a room with only the 16 items or in a full room. Pezdek et al. found that, while adhering closely to the experimental design of Brewer and Treyens, inconsistent objects enjoyed better recall and recognition. As a result they argued that the Brewer and Treyens findings were due to the relatively few inconsistent objects in their office scene.

To evaluate the consistency effect in naturalistic scenes we evaluated object consistency across the 25 images in our scene study. In our experiment, we had five images for each of the five scene types. Objects were scored for the occurrence frequency within a scene type. For example, all of the five kitchen images contained a stove, leading to a consistency score of 5 (high consistency). Across the five kitchen images, there was only one image that contained a small wooden sailboat leading to a consistency score of 1 (low consistency). In this example, the wooden sailboat is a novel object for a kitchen scene and was not generated as a response by any of the participants in the prior knowledge study. Figure 5 shows the (correct) recall probability given this consistency score across all scene

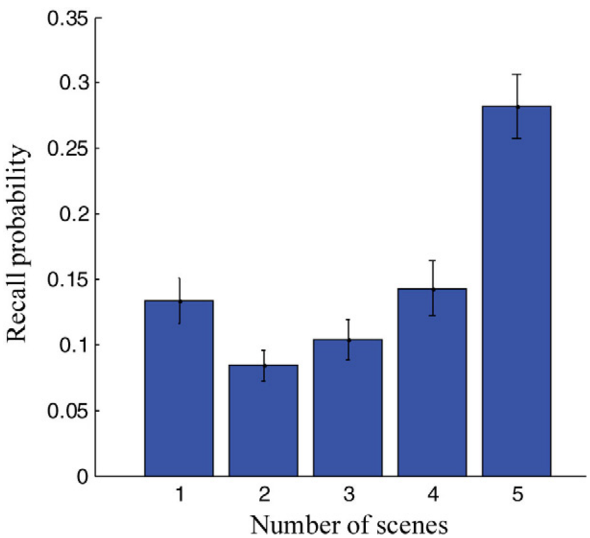

Figure 5 The effect of object consistency on recall probability. Object consistency for an object is measured by the number of scenes the object is present in. Objects occurring in only one scene are considered novel whereas objects occurring in five scenes are considered consistent with the scene type.(For color version of this figure, the reader is referred to the web version of this book.)

types and images in the memory experiment. We found that *both* consistent and novel objects enjoy better recall. As discussed earlier, both effects have been demonstrated in the literature, but one effect is usually elicited at the cost of the other. When memory is studied in situations where the natural environment has been manipulated—as was the case in both the Brewer and Treyens and Pezdek et al. studies—only the effects of interest emerge. This is likely to be due to choices in the particular manipulation of the stimulus material. For example, Pezdek et al. in effect binarized their stimuli to be either consistent or inconsistent, but they did not vary in degree of consistency, while consistency for Brewer and Treyens was continuous. Pezdek et al. also included the same number of inconsistent and consistent objects. By definition however, there are only a few inconsistent objects available in naturalistic scenes and when the stimuli are unaltered—as was the case in our study—both the effect of consistency and the effect of inconsistency emerge naturally.

Obviously, in naturalistic environments the content of the scene is outside the control of the experimenter. Better recall for consistent objects is due to the fact that in a natural scene high prior expectation object are by definition likely to be present in the scene. Better recall for inconsistent objects might be due to the fact that prior expectations allow for very fast gist extraction of a scene (Potter, Staub, Rado, & O'Connor, 2002), making inconsistent objects more salient. In this way prior knowledge plays an important role in recall both for items for which one might have high prior expectations as well as for unexpected items.

7. The Contribution of Prior Knowledge

The difficulty in evaluating the relative contribution of prior knowledge on recall for events is that we cannot take out the contribution of prior knowledge when recalling events. We can however, study prior knowledge in the absence of episodic memory. One possibility for studying the impact of prior knowledge without episodic memory is to ask people to make guesses about studied scenes or events even when they have not been exposed at all to any of these scenes or events, forcing them to generate responses based on prior knowledge alone. This corresponds to a memory experiment where the stimulus was presented for 0 s study time. Even though we did not actually run this in the memory experiment, we can consider the responses from the prior knowledge norming experiment as reasonable guesses to the objects that might be present in a particular scene. Figure 6 shows the cumulative accuracy in the memory experiment as a function of output position. The cumulative accuracy is the proportion of correct responses from the first *n* recalled objects. For example, suppose

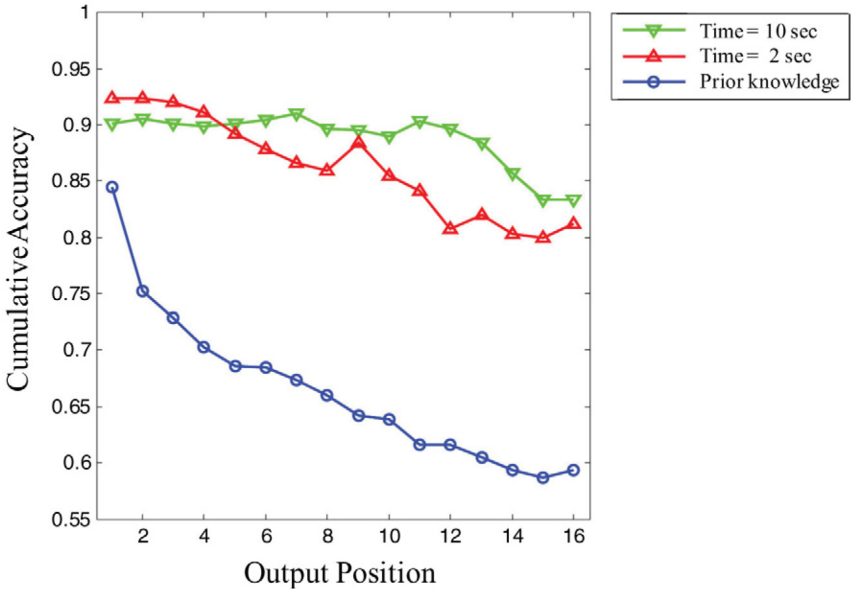

Figure 6 Cumulative accuracy as a function of study time and output position. The prior knowledge condition shows the performance when one treats the responses from the prior knowledge experiment as responses in the memory experiment. (For color version of this figure, the reader is referred to the web version of this book.)

a participant recalled 10 objects and for the first 4 objects, there were 3 correct responses (and 1 intrusion), the cumulative accuracy for output position 4 is 0.75. Figure 6 shows the cumulative accuracy in three different conditions: the 2 and 10 s study time conditions, and the prior knowledge condition where we treated the prior knowledge responses for each scene type as memory responses for the image (for the same type), preserving the order of the responses. Performance in this condition is fairly high. The first item guessed in the prior knowledge experiment leads to 85% accuracy in the memory experiment, even though the response is not based on any episodic information of the presented scenes. For later responses, accuracy does decrease but cumulative accuracy is still higher than 55% even after guessing 16 items. The difference between the performance from prior knowledge and actual recall reveals the contribution of episodic memory. Our finding that prior knowledge can provide a significant contribution to recall is consistent with some previous studies. Anderson and Pichert (1978), for example, showed that active schemas (i.e., prior knowledge) aid recall, and Brewer and Treyens point out that prior knowledge and recall are naturally correlated. What is unexpected is the very large baseline contribution of prior knowledge alone. One would expect recall to be based predominantly on episodic memory traces, not on prior knowledge. Our results demonstrate that general knowledge of scenes can greatly contribute to the accuracy of recalling objects from natural scenes.

Figure 6 also shows that cumulative accuracy decreases as a function of output positions when the scene was actually presented in the memory experiment. Therefore, intrusions are introduced later in recall, a finding compatible with results from the verbal memory domain (Roediger & McDermott, 1995). Cumulative accuracy was highest for the short study time condition for the first five output positions. After the sixth output position, the cumulative accuracy was best for the long study time conditions. Therefore, the somewhat counterintuitive finding here is that shorter study times do not necessarily lead to worse performance—the first few items remembered are *more* likely to be correct compared to a condition with longer study times (however, the *total* number of correct responses is greater with longer study times; for 2 and 10 s conditions, there were an average of 7 and 9 correct responses respectively per subjects per image).

We can explain this finding as an effect of the trade-off between episodic memory and semantic knowledge. For short study times, only a few objects might have been observed. Some of these objects can be encoded episodically without running into interference or capacity constraints. These few objects can subsequently be output with fairly high accuracy. On the other hand, if a scene is studied for a longer period, more objects overall are noticed and will need to be encoded. This longer list might not be encoded entirely by episodic means and part of the encoding might be based on generalized semantic knowledge. This will lead to lower accuracy for the first few items recalled but to higher accuracy at later output positions because of the enhanced semantic encoding.

In sum, prior knowledge can lead to good baseline performance in episodic memory tasks. When recalling objects from a kitchen that has never been seen before, recall can be reasonably good if the guesses are based on general knowledge of kitchen scenes (e.g., guesses such as "refrigerator," and "sink"). Of course, performance improves when actual episodic memories of the particular image can be retrieved, but the contribution of episodic memory is perhaps smaller than one would expect.

8. Towards Ecological Validity in Memory Research

Many researchers have considered arguments for and against ecological validity in memory research (e.g., Banaji & Crowder, 1989; Neisser, 1978). In our research, we investigate whether the use of naturalistic stimuli that is representative of the environment elicit findings that are different from those elicited with manipulated or constrained stimuli.

In studies with images of naturalistic scenes we have shown that this choice of stimuli can lead to very different conclusions about memory functions and the resulting errors. We found that there are relatively few intrusions in free recall for naturalistic stimuli. The intrusions are also more likely to be for objects with low prior expectations for a given scene rather than for objects with high prior expectations for a scene. This is in contrast to the findings of Brewer and Treyens, who showed that prior knowledge, in the form of high prior expectations, led to greater intrusion rates. Furthermore, we showed that both consistent and novel objects enjoy greater recall in naturalistic scenes. This synthesizes the findings of two papers in the literature that have independently argued for one or the other effect. Lastly, we showed that prior knowledge provides a high baseline performance in episodic memory. When scoring people's prior expectations as if they were performance in a memory study we found an unexpectedly high contribution from prior knowledge.

8.1. Some Arguments for Ecological Validity

What is it about the quality of naturalistic stimuli that might elicit these findings? Naturalistic stimuli that are representative of the environment hold information that might be inadvertently or deliberately absent from manipulated stimuli. High prior probability objects for example, are naturally present in the naturalistic environment. This makes intrusion errors, such as recalling books in an office where there are no books, less likely, because the high probability object is a priori likely to be present in a scene. In a naturalistic environment, the prior knowledge of the occurrence of objects in a given scene type can lead to effective guesses, even in the absence of any episodic information about the specific scene or event. Such guessing with prior knowledge can result in high accuracy and a low number of intrusions.

Adopting a strategy of guessing using prior knowledge might be a prudent strategy in recall, because it provides a natural bound on errors. Having prior knowledge about the natural size range of apples, for example, provides a bound on the possible sizes when recalling a particular studied apple. Using prior knowledge also reduces resources needed by the cognitive system at both encoding and retrieval. People have been shown to quickly and accurately extract the gist of a scene (Potter, Staub, Rado, & O'Connor, 2002). Having extracted the gist one can then use available cognitive resources for other purposes—given that the scene follows the prior expectations of the natural environment—and still maintain a high level of accuracy in recall. After the gist of a scene is extracted one can focus instead on unexpected objects in a scene. For example, Loftus and Mackworth (1978) showed that looking times were longer for nonschema related items, suggesting that having prior knowledge frees up cognitive capacity that can then be reallocated to objects that are inconstant with those prior expectations. In this way it is possible for *both* objects that are consistent with a scene *and* objects that are novel to a scene to enjoy high rates of recall.

Based on the findings presented here we would argue that using interesting, unaltered, and ecologically valid stimuli leads to findings that inform memory research in a broader context. While this results in some loss of experimental control we believe that it is worthwhile shifting memory research in this direction. We are not the first to argue for increased ecological validity in memory research. Most notably, Neisser (1978) argued that memory research should not only strive for greater ecological validity, but had in fact been marching down the wrong road ignoring the important questions of memory. Neisser advocated for a move out of the laboratory and in to the real world of everyday memory, to address more common questions of memory of greater interest to the layperson (e.g., "why can't I remember what I had for breakfast?"). We want to make it clear that we are not advocating such a move out of the laboratory. We would argue that ecological validity does not mean a complete departure from the laboratory setting. Neither are we suggesting that the focus of memory research should be to address more everyday questions of interest, or that is not necessary to investigate intuitive aspects of memory (as Neisser suggests). Our argument is about the stimulus used within the laboratory setting. We favor an approach where memory researchers adopt more ecologically valid stimulus material.

8.2. Considering Some Counterarguments

Several researchers have offered concerns regarding the application of the concept of ecological validity to areas of research such as education (Dunlosky, Bottiroli, & Hartwig, 2009), developmental psychology (Schmuckle, 2001), and executive function (Chaytor, Schmitter-Edgecomb, & Burr, 2006). In the area of memory, Banaji and Crowder (1989) published a vocal critique of the approach favored by Neisser. They argued that high ecological validity was not necessary, but rather that high generalizability was important in experimental design. Banaji and Crowder described a two-by-two array of high and low ecological validity and high and low generalizability. While they acknowledged that the ideal is both high ecological validity and high generalizability, they favored low ecological validity and high generalizability over high ecological validity and low generalizability. We would argue that ecological validity is not binary. Ecological validity can range over a broad spectrum, ranging from nonsense syllables to natural observation, where ecological validity can be pushed higher or lower on the range. While we favor ecological validity as an approach to the choice of stimulus material, we certainly do not want to suggest that the particular stimulus material we have adopted in our own research is on the high end of this range. The stimuli material we have adopted does however fit with the cell in Banaji and Crowder's array where high ecological validity meets high generalizability. The statistics of our stimuli more closely approximates the statistics of the natural environment. However, we did make choices regarding experimental design that sacrificed some ecological validity. For example, we tested subjects in a lab setting and not in the actual settings. The participants saw scenes in images, not the real environments, in contrast with the Brewer and Treyens office study where participants experienced the real environment. There are many ways of having more or less naturalistic stimuli. The balance for us was to select environments that were fairly natural, but at the same time could be measured and characterized.

9. Conclusion

Ecological validity is not uniquely aimed at memory research. Similar arguments have been made for research on cognition in general (Dhami, Hertwig, & Hoffrage, 2004), and has enjoyed much wider acceptance in visual perception (Gibson, 1966, 1979). The approach of ecological validity however, falls outside of the comfort zone of many researchers. Banaji and Crowder admit that high ecological validity is the ideal, but it is not an ideal that they themselves pursue. Ecological validity should not be pursued at the cost of generalizability, but the study of memory should also not be sacrificed to the comfort zone of standard practices. Memory is a relationship between the people doing the remembering and the world in which we operate. The cost of using rich, ecologically valid, unmanipulated stimuli is a decrease of experimental control, whereas the cost of exerting ultimate experimental control is studying memory in a vacuum—and potentially, erroneous conclusions about the function of memory. The challenge is how to find a good balance between ecological validity and experimental control.

REFERENCES

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*(1), 1–12.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist*, *44*(9), 1185–1193.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207–230.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.
Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, *20*, 718–721.
Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*, 217–227.
Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22.
Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*, 959–988.
Dunlosky, J., Bottiroli, S., & Hartwig, M. (2009). Sins committed in the name of ecological validity: A call for representative design in education science. In D. Hacker., J. Dunlosky, and A. Graesser, (Eds.), *Handbook of metacognition in education* NY: Psychology Press.
Ebbinghaus, H. (1964). Memory: A contribution to experimental psychology (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover Publications, Inc. (originally published, 1885).
Gibson, J. J. (1979). The ecological approach to visual perception. Boston, MA: Houghton-Mifflin.
Gibson, J. J. (1966). The senses considered as perceptual systems. Oxford, England: Houghton-Mifflin.
Hemmer, P., & Steyvers, M. (2009a). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*, 189–202.
Hemmer, P., & Steyvers, M. (2009b). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin and Review*, *16*, 80–87.
Hemmer, P., & Steyvers, M. (2009c). *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*, Integrating episodic and semantic information in memory for natural scenes.
Hemmer, P., Steyvers, M., & Miller, B. (2010). The wisdom of crowds with informative priors. In S. Ohlson, and R. Catrambone, (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1130–1135). Mahwah, NJ: Lawrence Erlbaum.
Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in establishing spatial location. *Psychological Review*, *98*, 352–376.
Huttenlocher, J., Hedges, L. V., & Prohaska, V. (1992). Memory for day of the week: A 5 + 2 day cycle. *Journal of Experimental Psychology: General*, *121*, 313–325.
Huttenlocher, J., Hedges, L. V., & Vevea, J. (2000). Why do categories affect stimulus judgments? *Journal of Experimental Psychology: General*, *129*, 220–241.
Kan, I. P., Alexander, M. P., & Verfaellie, M. (2008). Contribution of prior semantic knowledge to new episodic learning in amnesia. *Journal of Cognitive Neuroscience*, *21*, 938–944.
Kato, K., & Higashiyama, A. (1998). Estimation of height for persons in pictures. *Perception and Psychophysics*, *60*, 1318–1328.
Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in perception and memory. In D. S. McNamara, and J. G. Trafton, (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 407–413). Austin, TX: Cognitive Science Society.
Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490–517.
Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 565–572.

Neisser, U. (1978). Memory: What are the important questions? In M. M. Gruneberg., P. E. Morris, and R. N. Sykes, (Eds.), *Practical aspects of memory* (pp. 3–24). London: Academic Press.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory of real world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *5*, 587–595.

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.

Roediger III, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.

Russel, B. C., & Torralba, A. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*, 157–173.

Schmuckle, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*, 419–436.

von Restorff, H. (1933). Uber die Wirkung von Bereichsbildungen im Spurenfeld [the effects of field formation in the trace field]. *Psychologische Forschung*, *18*, 299–342.