UNIVERSITY OF CALIFORNIA, IRVINE

The Influence of Prior Knowledge on Reconstructive Memory

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Psychology

by

Pernille Hemmer

Dissertation Committee:
Professor Mark Steyvers, Chair
Professor Michael Lee
Professor Elizabeth Loftus

2011

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Professor Mark Steyvers, who demonstrates an enthusiasm and creativity in research that has served as an inspiration for my work. Without his guidance and persistent help this dissertation would not have been possible.

I would like to thank my committee members. Professor Michael D. Lee has been an invaluable support in all matters science, as well as a dear friend. Professor Elizabeth Loftus, whose work demonstrated to me that research has an important reach outside of academia and who is a role model for women in science.

Thank you to my fellow MAD Lab members, especially Miles, Brent and Shunan, who have been good friends, collaborators and travel companions. My sincerest appreciation also goes out to a small army of research assistants without whom my work would not be what it is. Most notable among them Jenny Shi, whose research exploits are a tour de force and who opened my eyes to animated Taiwanese news.

I would also like to thank my parents. Til mor, tak for at du altid har troet på mig. Til far, du er altid i mine tanker.

Finally, I deeply thank John who supported and encouraged me throughout my never-ending quest for higher education. I love you!

Chapter 1, in full, is a reprint of the material as it appears in Integrating Episodic Memories and Prior Knowledge at Multiple Levels of Abstraction in *Psychonomic Bulletin & Review*, 16 (1), 80-87. Hemmer, P. & Steyvers, M. (2009). The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in A Bayesian Account of Reconstructive Memory in *Topics in Cognitive Science, 1*, 189-202. Hemmer, P. & Steyvers, M. (2009). The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in Integrating Episodic and Semantic Information in Memory for Natural Scenes in N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp.1557-1562). Austin, TX: Cognitive Science Society. Hemmer, P. & Steyvers, M. (2009). The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in The Wisdom of Crowds with Informative Priors in S. Ohlsson & R.Catrambone (Eds.), *Proceedings of the 32st Annual Conference of the Cognitive Science Society* (pp.1130-1135). Austin, TX: Cognitive Science Society. Hemmer, P., Steyvers, M. & Miller, B. (2010). The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, has been submitted for publication and the material of it may appear in The Influence of Real World Prior Knowledge on Episodic Memory in *Memory & Cognition*. Hemmer, P., Shi, J. & Steyvers, M. (2011). The dissertation author was the primary investigator and author of this paper.

Financial support was provided by a Graduate Deans Dissertation Fellowship at the University of California, Irvine.

CURRICULUM VITAE

EDUCATION

2011	PhD, University of California**,** Irvine
2009	M.A. Psychology, University of California, Irvine
2006	B.A. Psychology, University of California, Irvine
	Summa cum Laude

HONORS AND AWARDS

2010	Recipient, Graduate Dean's Dissertation Fellowship
	Recipient, Society for Mathematical Psychology Student Travel Award
	Recipient, Jean-Claude Falmagne Research Award
	Recipient, Spring Social Science Graduate Fellowship
2009	Recipient, John I. Yellott Scholar Award
	Recipient, Cognitive Science Society NSF Student Travel Award
	Recipient, Society for Mathematical Psychology Student Travel Award
2008	Recipient, *Cognitive Science* best paper on modeling higher-level cognition
2006	Recipient, School of Social Sciences Graduate Fellowship
	Recipient, Order of Merit
2005-2006	Member, Student Editorial Board, the UCI Undergraduate Research Journal
	Member, Deans Honor List
2004-2006	Member, Honors in Psychology Program
2005	Recipient, Undergraduate Research Opportunities Fellowship
	Recipient, Summer Undergraduate Research Opportunities Fellowship
	President, Psi Chi psychology honors society

Appointments

2007-	Graduate Liaison to the Department of Cognitive Sciences
2009-	Student Member, Curriculum Committee, Department of Cognitive

PROFESSIONAL AFFILIATIONS

2009-	Society for Mathematical Psychology
2008-	Cognitive Science Society
2008-	Association for Psychological Science

Ad-hoc reviewing

Psychonomic Bulletin and Review
Memory & Cognition
Cognitive Science

RESEARCH

2010-	Estimating Individual Differences in a rational Model of Memory
	Primary researcher
	Faculty mentor Professor Mark Steyvers, Dept. of Cognitive Sciences
2010-	The influence of Prior Knowledge on Recall for Height
	Primary researcher
	Faculty mentor Professor Mark Steyvers, Dept. of Cognitive Sciences
2009	Assessing Episodic and Semantic Components in Serial Recall
	Primary researcher

	Faculty mentor Professor Mark Steyvers, Dept. of Cognitive Sciences
2009	Integrating Episodic and Semantic Information in Memory for Natural Scenes Primary researcher Faculty mentor Professor Mark Steyvers, Dept. of Cognitive Sciences
2009-	The Wisdom of Crowds in Rank Ordering Tasks Co-researcher Faculty Mentor Professor Mark Steyvers, Dept. of Cognitive Sciences
2006-2008	Integrating Episodic Memories and Prior Knowledge at Multiple Levels of Abstraction Primary researcher Faculty mentor Professor Mark Steyvers, Dept. of Cognitive Sciences
2005	Decision Making in recognition Tasks in Alternating Environments Primary researcher Faculty mentors Professor Scott Brown and Professor Mark Steyvers Dept. of Cognitive Sciences

Undergraduate students supervised

2010-	Mindy DeYoung (research assistant); Daniela Amado (research assistant).
2009 -	Jenny Shi (mentored for SURP and UROP fellowships).
2008-2009	Stanton Mack (research assistant); Christina Maranhao (research assistant); Erin Horowit (research assistant).

PUBLICATION LIST

Journal Articles

Hemmer, P. & Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. *Topics in Cognitive Science, 1*, 189-202.

Hemmer, P., Steyvers, M. (2009). Integrating Episodic Memories and Prior Knowledge at Multiple Levels of Abstraction. *Psychonomic Bulletin & Review, 16(1), 80-87.*

Brown, S.D., Steyvers, M., & Hemmer, P. (2007) Modeling Experimentally induced Strategy Shifts. *Psychological Science, 18*, 40-45.

Hemmer, P., Shi, J., & Steyvers, M. (Submitted). The influence of Real World Knowledge on Recall for Height.

Papers in refereed conference proceedings

Hemmer, P., Steyvers, M, & Miller, B. (2010). The Wisdom of Crowds with Informative Priors. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Steyvers, M., Lee, M.D., Miller, B., & Hemmer, P. (2009). The Wisdom of Crowds in the Recollection of Order Information. In J. Lafferty, C. Williams (Eds.) *Advances in Neural Information Processing Systems, 23*. MIT Press.

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* (pp. 1557-1562). Austin, TX: Cognitive Science Society.

Miller, B., Hemmer, P. Steyvers, M. & Lee, M.D. (2009) The Wisdom of Crowds in Rank Ordering Tasks. *Proceedings of the 9th International Conference of Cognitive Modeling.*

Hemmer, P. & Steyvers, M. (2008). A Bayesian Account of Reconstructive Memory. In V. Sloutsky, B. Love, and K. McRae (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum
Prize for computational modeling of higher-level cognition

Published Abstracts

Hemmer, P., Shi, J. & Steyvers, M. (2010). The influence of Real World Knowledge on Recall for Height [Abstract]. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (p. 2402). Austin, TX: Cognitive Science Society.

PRESENTATIONS AT CONFERENCES AND MEETINGS

2010	*Annual Meeting of the Psychonomic Society*, St. Louis.
	Annual Conference of the Cognitive Science Society, Portland, Or
	Annual Meeting of the Society for Mathematical Psychology, Portland, Or
	Annual Summer Interdisciplinary Conference, Bend, Or
	Context and Episodic Memory Symposium, Philadelphia
2009	*Annual Meeting of the Psychonomic Society*, Boston.
	Annual Meeting of the Society for Mathematical Psychology, Amsterdam
	Annual Conference of the Cognitive Science Society, Amsterdam
	Annual Summer Interdisciplinary Conference, Valle d'Aosta, Italy
2008	*Annual Meeting of the Cognitive Science Society, Washington DC*
	Annual Summer Interdisciplinary Conference, Madonna Di Campiglio, Italy
	Context and Episodic Memory Symposium, Tampa, Florida
2007	*Annual Summer Interdisciplinary Conference*, Kalymnos, Greece.
	Annual meeting of the Society for Mathematical Psychology, Costa Mesa, Ca
2005	*Annual Meeting of the Psychonomic Society*, Toronto.

Invited talks

2011	*Rutgers University*. Colloquium
2010	*University of London, Birkbeck.* Cognitive Science Seminar Series

TEACHING EXPERIENCE

University of California, Irvine
Teaching Assistant, Psychology Fundamentals, Summer 2010
Teaching Assistant, History of Psychology, Summer 2010
Teaching Assistant, Honors Experimental Psychology, Spring, Fall 2009, Winter 2011
Teaching Assistant, Advanced Experimental Psychology, Winter, Spring 2008, Winter 2009
Teaching Assistant, Psychology Research methods, Fall 2008
Teaching Assistant, Upper division Applied Statistics, Fall 2005, Winter, Spring 2006

RELATED PROFESSIONAL EXPERIENCE

University of California, Irvine Memory and Decision Making Lab
Research Assistant, Fall 2006, Winter, Spring, Fall 2007, Winter 2010
Starbucks Coffee Company
Store Manager, 2001-2004
IKEA
Business Manager, 1996-2001

ABSTRACT OF THE DISSERTATION

The Influence of Prior Knowledge on Reconstructive Memory

By

Pernille Hemmer

Doctor of Philosophy in Psychology

University of California, Irvine, 2011

Professor Mark Steyvers, Chair

Prior knowledge and expectations about events exert a strong influence on reconstructive memory. In this thesis I investigate the interaction between memory and prior knowledge in four different task domains – recalling the size of fruits and vegetables, objects in scenes, the order of events and the height of people. In contrast to previous work, I assess peoples' prior knowledge for each domain experimentally. I then assessed the influence of prior knowledge in a series of memory experiments in which the actual stimuli shown during a study phase is compared with the stimuli reconstructed during the test phase.

I present evidence of the hierarchical influences of prior knowledge in the height and fruit-and -vegetable tasks. These tasks show that the reconstruction of stimuli for which participants have finer grained knowledge – e.g., the size of apples and strawberries – is influenced toward the specific prior for that type of object, but stimuli for which participants have more general knowledge – e.g., abstract shapes – are influenced toward the overall category size.

From the natural scene and event order tasks I present evidence that prior knowledge can also lead to good baseline performance in recall. For example, I ask participants to order scrambled event sequences without ever viewing the sequence during study. This is used as a measure of how well people can perform in reconstructing events based on prior knowledge alone. I find that for some stereotyped event sequences, peoples' performance using only prior knowledge is almost as good as if they had seen the actual sequence of events.

Overall the results demonstrate that general knowledge can greatly contribute to the accuracy of recall and that prior knowledge can improve memory in most situations that occur naturally.

Introduction

Prior knowledge and expectations about events are known to influence episodic memory (e.g. as described in the seminal studies by Bartlett, 1932). In many situations, episodic retrieval involves a component of reconstruction where the information from episodic memory is combined with prior knowledge in the retrieval process. For example, suppose you have to recall events from a recent visit to a coffee shop. You might not only utilize the episodic information related to that specific coffee shop visit, but also general knowledge and experiences accumulated over many coffee shop visits. You might infer that you ordered a tall coffee at your last visit, not because you have detailed explicit memories of this event, but because you typically order tall coffees. In this coffee shop scenario, there are many aspects that might be remembered in similar fashion, such as recalling the height of the person that stood behind you based on remembering the gender of the person alone, the objects that were present in the coffee shop, or the sequence of events that transpired in the coffee shop. This scenario illustrates that 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.

Episodic memory and prior knowledge can also interact during the encoding process. For example, you might remember that you ordered a very large coffee because it deviates from your typical behavior. Because you rarely order such coffees, this deviation from the norm leads to heightened attention and event specific encodings in episodic memory.

The goal of the studies reported in this paper is to better understand the interaction between prior knowledge and episodic memory. How do we integrate noisy and incomplete information stored in episodic memory with prior knowledge? All the studies have three overarching themes. First, I assess the prior knowledge that people bring to the memory task and use some novel methods to extract and characterize this knowledge. Second, I utilize memory tasks that are ecologically valid. The stimulus materials to be remembered have the same statistical regularities as can be found in the natural environment. Third, the studies are designed to demonstrate the potential benefit of prior knowledge. There are many demonstrations in the literature that show that prior knowledge can lead to errors. While such errors might be an inevitable result of utilizing prior knowledge in episodic memory tasks, I work to show that the *overall* accuracy in episodic memory improves as a result of prior knowledge. Therefore, in a hypothetical situation where we would have control over our retrieval process and we decided not to utilize prior knowledge, I would expect that this would *decrease* the overall performance of our memory system.

The results of the studies provide converging evidence that prior knowledge interacts with episodic memory at multiple levels of abstraction, that prior knowledge can improve overall accuracy in reconstruction from memory, and that items associated with strong prior expectations are not the primary source of errors.

In the first four chapters I will present published research of three studies. Chapter 1 is a study, on the influences of prior knowledge in recall for size of objects, which appeared in *Psychonomic Bulletin & Review*. Chapter 2 presents a Bayesian analysis of the empirical study presented in chapter 1. This paper originally appeared in *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. The paper was awarded the prize for best paper on computational modeling of higher level cognition. It was later reprinted in the invitational first issue of *Topics in Cognitive Science* (this is the version presented here). Chapter 3 presents a study on the influence of prior knowledge on recall for objects in scenes. This paper appeared in the *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Chapter 4 presents a study on the influence of prior knowledge on recall for temporal events. This paper appeared in the *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. These papers are presented exactly as they appeared in the journals, except for the reformatting to fit this document, and the collating of all references at the end of this document.

The last chapter presents a study on the influence of prior knowledge on recall for the height of people. This is an unpublished manuscript that has been submitted to *Memory & Cognition* for consideration. Together this work seeks to develop a comprehensive understanding of the interaction between knowledge and memory, and toward the development of a theoretical and modeling framework of reconstructive memory.

Chapter 1

Integrating episodic memories and prior knowledge at multiple levels of abstraction

Pernille Hemmer & Mark Steyvers (2009). *Psychonomic Bulletin & Review, 16 (1), 80-87*

Abstract

Prior knowledge can have a large influence on recall when the memory for the original event is error prone or incomplete. We investigated the interaction between memory and prior knowledge in a recall task involving natural objects such as fruits and vegetables. We first quantified prior knowledge for the sizes of objects in a norming experiment. We then assessed the influence of prior knowledge in a memory experiment in which we compared the actual size of objects shown during a study phase with the reconstructed size of an object during the test phase. Recall was biased both by the mean size of the specific object studied and by the mean size of all objects in the category. This result suggests that the influence of prior knowledge can come from multiple, hierarchically related levels of representation, such as the object-category and superordinate-category levels.

Introduction

Reconstructing events from memory involves the coordination of multiple sources of information. Recall of past events, such as the time we went to work or the size of the coffee we ordered last week, might be based partially on vague recollections of the events themselves but also on prior knowledge; perhaps we usually go to work at a specific time and tend to order coffee of a specific size. Such knowledge might provide useful cues when recalling past events. Bartlett’s (1932) seminal research on reconstructive memory initiated many research paradigms that demonstrated how cultural and social norms, as well as cognitive expectations, influence our recall of past events. For example, height judgments can be biased by gender (Biernat, 1993), and face recognition can be influenced by ethnicity (Rehnman & Herlitz, 2006). Sometimes the expectations based on prior knowledge can lead to systematic errors and intrusions in recall. An office without books can later be remembered as having books, presumably because most offices are expected to have books (Brewer & Treyens, 1981). Similarly, words can be falsely recalled from a list, especially when the list is structured to have strong associations to a word not present on the list (Roediger & McDermott, 1995).

On the other hand, prior knowledge can also be an aid to memory. Recall for abstract information typically improves when additional meaningful information is available (Bartlett, 1932; Bower, Karlin, & Dueck, 1975; Bransford & Johnson, 1972). Prior knowledge can also be advantageous in recall when the episodic memories for the original event are noisy or incomplete. Huttenlocher and colleagues (Crawford, Huttenlocher, & Engebretson, 2000; Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000) presented a model of category effects in which reconstruction from memory is based on a weighted average of episodic memory traces and prior knowledge in the form of category information. By combining both episodic and category information in recall, large fluctuations in reconstructions due to noisy episodic memories can be prevented. The result of this weighted average is that reconstructions are systematically biased toward the category center. Such systematic biases have also been found in memory psychophysics experiments that reveal a compression of estimated stimulus attributes when those attributes are recalled from memory (Kerst & Howard, 1978; Moyer, Bradley, Sorensen, Whiting, & Mansfield, 1978).

Previous studies (e.g., Huttenlocher et al., 1991; Huttenlocher et al., 2000) focused on categorization, using relatively short-term memory tasks in which participants were tested immediately after stimulus exposure. With such short lags, retrieval from episodic memory is less error prone, and it might be more difficult to observe the influence of prior knowledge. Previous studies have also relied on estimation of relatively artificial stimuli, such as the locations of dots placed in a circle, the lengths of lines, or the sizes of blue and red circles (e.g., Huttenlocher et al., 1991; Huttenlocher et al., 2000; Sailor & Antoine, 2005). Participants are unlikely to have strong preexperimental knowledge for such stimuli, and instead, the prior knowledge is created during the course of the experiment by training participants on different stimulus distributions. In addition, some researchers have argued that any central bias can simply be explained by sequential effects (Sailor & Antoine, 2005). Because the smallest stimuli are always preceded by a larger stimulus, and vice versa, temporal contiguity can result in a central bias, and no explanation in terms of prior knowledge is needed.

In this research, we investigated the interaction between prior knowledge and recall, using an experimental paradigm that differs in three important respects from previous studies. First, our memory tasks emphasized memory on a longer time scale, which we expected to lead to a decrease in the accuracy of the episodic memory representations and a concomitant greater reliance on prior knowledge.

This allowed us to better investigate the influence of prior knowledge on recall and to emphasize the memory process. Second, unlike in previous studies, we employed naturalistic stimuli, such as fruits and vegetables, for which we expected relatively well-established preexperimental knowledge. We also expected that such naturalistic stimuli would be associated with more structured knowledge representations in which knowledge could be described at several levels of abstraction. In our experiments, we tested memory for one-dimensional stimulus attributes— namely, the size of an object. Prior knowledge for such attributes might exist not only at the superordinate level (e.g., "I expect fruits to be roughly of this size") but also at the object level (e.g., "I expect an apple to be of this size"). Therefore, the inclusion of naturalistic stimuli allowed us to study the influence of prior knowledge at multiple levels of abstraction. Our experiments were furthermore designed to evaluate systematic deviations that are a result of the influences of prior knowledge and not an artifact of sequential dependencies.

In two experiments, we investigated the interaction between prior knowledge and episodic memory. In Experiment 1, we collected norms for the sizes of natural objects, such as fruits and vegetables. In this norming experiment, we measured the expected sizes of common natural objects. Participants were asked to make judgments both about the mean expected size and about the minimum and maximum expected sizes of the target objects. In Experiment 2, we measured performance in a reconstructive memory task involving the normed stimuli.

Experiment 1

Method

Participants

Eighteen undergraduate students at the University of California, Irvine participated in the experiment. They were compensated with course credit.

Materials and Procedure

We sampled 24 high-resolution color images of fruits and vegetables from a large image database (PhotoObject.net). All of the images were photographed against a white background. The fruit and vegetable images had average sizes of 1,126 X 1,078 and 1,175 X 878 pixels, respectively. These images and a comparison image were presented on two computer screens. Each fruit or vegetable target image was presented on the right screen, along with a slider used to manipulate the size of the object. The comparison image, containing a display of a plate and utensils, was presented on the left screen. Participants were asked to make the following three judgments relative to the comparison image: "What is the average size of an object like this?" "What is the smallest size of an object like this?" and "What is the largest size of an object like this?" Participants resized the image using the slider. Responses were measured on a scale from 0 to 1, where 0 corresponded to an object scaled to less than 1 pixel on the screen and 1 corresponded to the maximum position of the slider, at which the target object fills either the entire height or width of the screen. The object remained on the screen until the participant was satisfied with the current size judgment and clicked a button to continue. Participants were shown real-life versions of the comparison object (a set of utensils) to ensure agreement on the common size of the object. The images were blocked by category (fruits and vegetables), and the 24 images within each category were presented in random order. The initial size for each image was chosen at random from the following sizes on the slider scale: .2, .4, .6, and .8 (proportions of the maximum size of the image).

Results

Figure 1 shows the mean ratings for 24 objects from the fruit and vegetable categories, ordered by the mean ratings for the average size.[1] The ordering of sizes reflects the intuitive notions that the participants had about fruits and vegetables. For example, in the fruit category, the average size judgment for a raspberry was smaller than that for a strawberry, which was smaller than that for an apple, all of which in turn were smaller than that for a pineapple. The range of the size judgments also increased with the magnitude of the objects, such that the range for the raspberry was the smallest, and the range for the pineapple was the largest. The participants expressed a reasonable degree of agreement in their judgments. For the *average* judgment, the mean pairwise rater correlation was .65, with a 90% confidence interval of .48 and .81

Experiment 2

In this experiment, we assessed how prior knowledge for the sizes of objects influences the recalled size of a particular object studied in the context of the experiment. We tested recall for the natural objects that were normed in Experiment 1, as well as for a set of artificial shapes. We predicted that the effect of prior knowledge for fruits and vegetables would occur at two levels of abstraction: the object-category and superordinate-category levels. We hypothesized that recall would be biased toward the mean of the distribution associated with all size variations of the particular object (e.g., an apple) and also toward the mean of the size distribution associated with all objects within the object's superordinate category (e.g., fruits). Therefore, we predicted that an object that is studied at a relatively small or large size relative to the mean of the object would be over- or underestimated, respectively, at test. Similarly, we predicted that an

---

[1] Data from 4 participants were discarded because these participants made *large* judgments that were smaller than their *small* judgments (or vice versa), indicating that they did not understand the task.

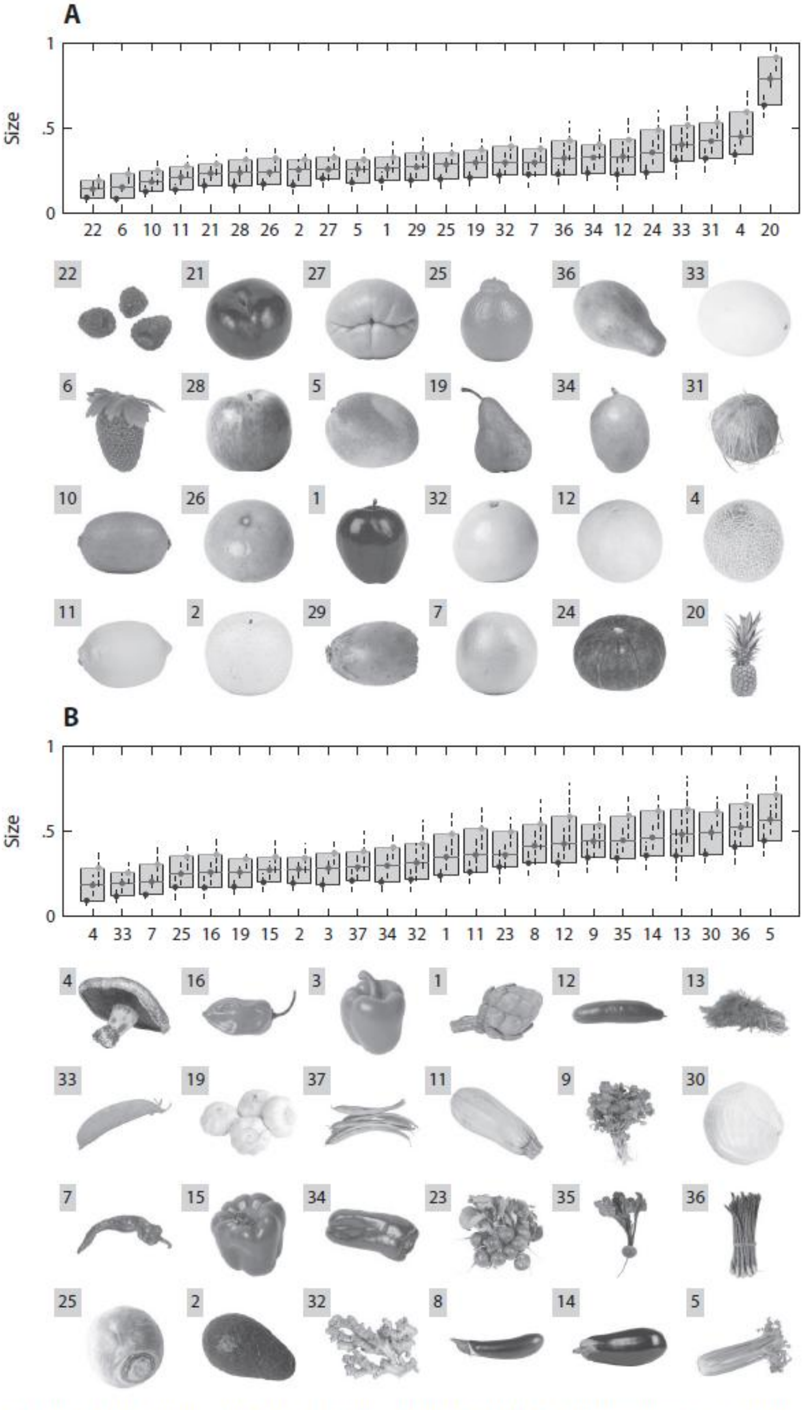


Figure 1.1 Mean rating results from Experiment 1 for the fruit (A) and vegetable (B) objects. The objects are in reading order by the mean of the *average* rating. The bar graph shows the range of size judgments for individual objects. For each bar, the means of the small, average, and large sizes are indicated by the bottom, middle, and top lines, respectively. The vertical lines correspond to the 25%–75% confidence interval across participants.

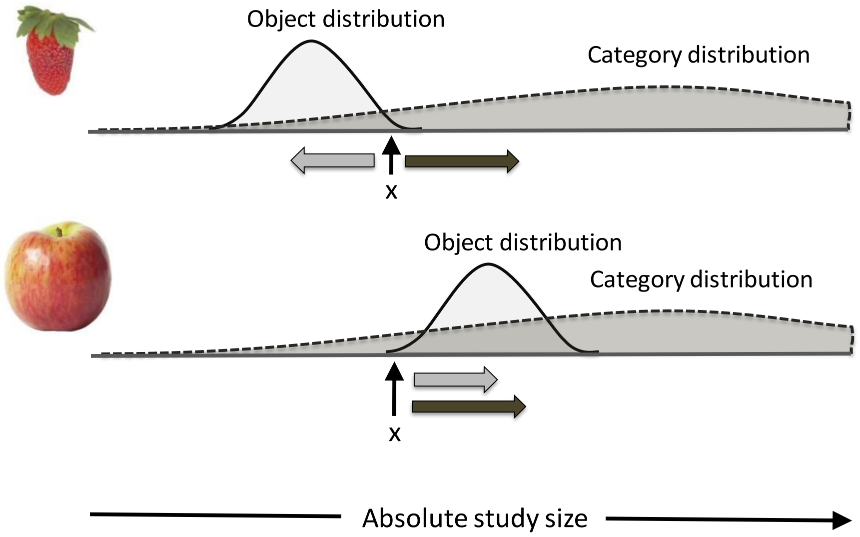


Figure 1.2 Predicted influences of prior knowledge at the object and superordinate levels. The distributions illustrate the hypothetical range of sizes associated with the fruit category and two particular objects: a strawberry and an apple. If the two objects are shown at the same study size, marked by "X," the strawberry and apple are presented at a very large and very small size, respectively, relative to their object distributions. This leads to object bias in opposite directions, toward the center of the respective object distributions, as shown by the light-colored arrows. In contrast, the superordinate bias operates in the same direction, toward the center of the fruit distribution, as shown by the darkcolored arrows.

object that is studied at a relatively small or large size relative to the mean of the superordinate category distribution would be over- or underestimated, respectively, at test. These predictions are consistent with the category effects found by Huttenlocher et al. (1991; Huttenlocher et al., smaller at reconstruction. Figure 2 illustrates how the contributions of prior knowledge at the object- and superordinate-knowledge levels can be teased apart. It illustrates the predicted recall for two example objects—a strawberry and an apple—shown at the same study size. An object-level bias would lead to an underestimation for the strawberry and an overestimation for the apple toward the centers of their respective object size distributions. A superordinate-level bias would lead to an overestimation for both the apple and the strawberry. Because these effects might operate simultaneously, it is unclear what the combined result would be of the object level and superordinate-level biases. However, the key prediction was that the objects that are studied at the same absolute study size can be *differentially* biased when the objects are associated with different object-level prior knowledge. This prediction also excludes the possibility that the bias is due to sequential effects or edge effects, whereby participants are reluctant to use the edges of the response scale.

In addition to the natural objects, we also tested recall for artificial shapes that have no natural size scale (see Figure 3) and presented these at the same sizes as the vegetable stimuli. The inclusion of such shapes allowed additional comparisons of the relative influences of prior knowledge for artificial and naturalistic stimuli. We expected that participants would not have any strong preexperimental expectations about size at either the object or the superordinate level. However, it was possible for participants to gain knowledge about the overall sizes of artificial shapes during the course of the experiment. Therefore, we hypothesized that recall would be biased toward the mean within the overall size distribution associated with the artificial shapes.

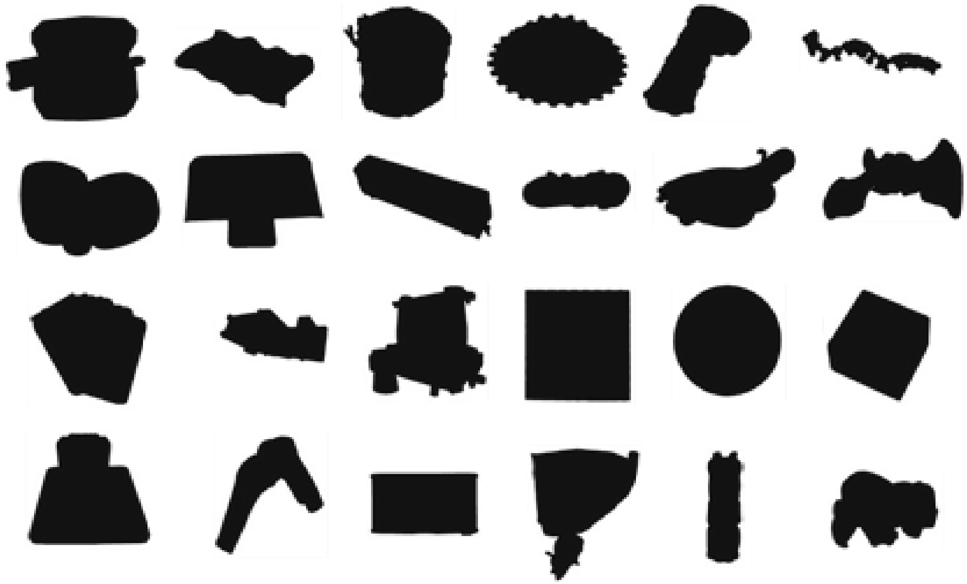


Figure 1.3. Artificial shapes used in Experiment 2.

Method

Participants

There were 25 participants, undergraduate students at the University of California, Irvine. The participants were not involved in the norming study in Experiment 1. They were compensated with course credit.

Materials and Procedure

We used the 24 objects from the fruit and vegetable categories in Experiment 1. We also created a category of 24 artificial shapes, which are shown in Figure 3. These images were created by drawing outlines of objects and filling the interiors with a blue color. We employed a recall paradigm analogous to continuous recognition, in which study and test events are randomly interleaved. The images were blocked by category (fruits, vegetables, and shapes), and each of the 24 images within each category was studied and tested three times, for a total of 144 trials per block. Trials were presented in random order but were constrained such that all test images had been previously seen as a study image, and there was always an intervening test trial between each repeated study trial. Furthermore, the study size for a given image remained constant across the three repetitions. Each study image was presented for 2 sec at the center of the right screen. The same comparison image that was used in Experiment 1 remained on the left screen at all times. Each test image was presented on the right screen, and the participants responded by clicking on one of two buttons on the screen: “Remember” or “Don’t Remember.” After this recognition judgment, the participants were then asked to scale the size of the test object using the mouse to move the slider presented on the right screen. The following question appeared with the slider: “What was the size of this object, compared to the object on the left, when you saw it at study?” Responses were measured on a scale of 0 to 1. The participants were given clear verbal instructions in order to ensure that they understood the task and were shown real-life versions of the comparison object (a set of utensils) in order to ensure agreement on the frame of reference.

On study trials, the size of a fruit or vegetable object was selected from the ranges shown in Figure 1. For each object, the range extended from the mean of the small judgments to the mean of the large judgments. This range was divided into eight bins, separately for each object. Five percent of study sizes were drawn from Bins 1 and 8 (smallest and largest), 10% from each of Bins 2 and 7, 15% from each of Bins 3 and 6, and 20% from each of Bins 4 and 5 (middle bins). This procedure ensured that the majority of the objects were presented at a size close to the mean size of the object and that few objects were shown at a size close to the smallest and largest acceptable sizes for that object. It also guaranteed that objects would never be presented outside their acceptable range. For the artificial shapes, the study sizes were sampled from the same distributions as for the vegetables. Each artificial shape was yoked to a vegetable object, such that the study size was the same as for the paired vegetable. On test trials, the images were shown randomly at one of four sizes on the slider scale: .2, .4, .6, and .8. The slider was initiated at the corresponding location.

Results

To investigate the effect of prior knowledge at the object-category and superordinate-category levels, we measured recall error as the difference between the recalled size and the studied size. We restricted the analysis to hits only (i.e., *remember* responses). For visual clarity, we divided the study objects in the vegetable and fruit categories into four classes on the basis of the study size relative to the size ranges obtained in the norming study. We named the four classes *very small*, *small*, *large*, and *very large*. These sizes do not refer to the absolute size at

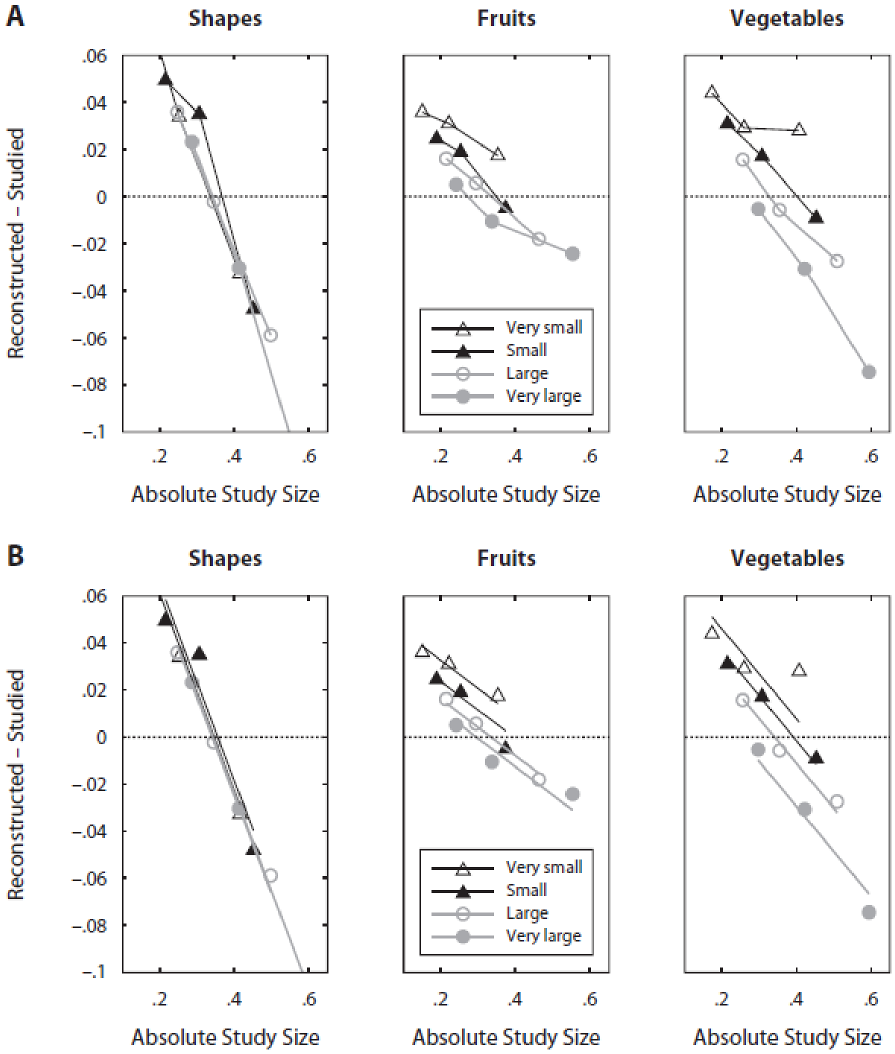


Figure 1.4. Recall biases as a function of absolute and relative study size. (A) Experimental data. Results are grouped by superordinate category and relative object size. (B) Linear regression model fits. Negative slopes are consistent with biases toward the center of the superordinate category range. Intercept differences are consistent with object-level knowledge effects.

which an object was studied but rather to the size relative to the distribution at the object level (e.g., the example in Figure 2 illustrates two objects presented at the same absolute study size but at different relative study sizes—a very small apple and a very large strawberry). Figure 4A shows the recall errors as a function of absolute study size for each superordinate category. The lines in the figure have two important aspects: slope and intercept. A negative slope indicates a bias to the center of the category range, such that small objects are overestimated (positive bias) and large objects are underestimated (negative bias). The difference between intercepts tests our key prediction that objects studied at the same absolute study size can be differentially biased, depending on prior knowledge at the object level. An intercept difference indicates a bias to the center of the object range, such that objects that are presented at a size that is small relative to the object size distribution have a greater overestimation error than do objects presented at a relatively large size. In order to statistically assess these effects, a linear regression model was fitted separately at the individual subject level. The regression model contained five parameters for each category: four intercept parameters, corresponding to the four relative object classes, and a single slope parameter (see Figure 4B for fits of this regression model). Table 1 shows the mean estimated slopes and intercepts across superordinate categories and relative object sizes. Separate *t* tests revealed that the slope for each category was significantly smaller than 0 [fruits, *t*(24) = -4.714, *p* = .000; vegetables, *t*(24) = -5.657, *p* = .000; shapes, *t*(24) = -10.754, *p*= .000]. This result is consistent with a superordinate level influence of prior knowledge. An object-level effect is revealed by the intercept differences between relative object sizes. A repeated measures ANOVA for the shape category found no effect of intercept [*F*(3,72) = 0.453, *p* = .716].

Table 1. Mean Slopes and Intercepts by Category and Relative Object Size for Experiment 2

		Fruits		Vegetables		Shapes	
	Relative Object Size	*M*	*SD*	*M*	*SD*	*M*	*SD*
Slopes		-0.120	0.127	-0.191	0.169	-0.415	0.201
Intercepts	Very small	0.057	0.049	0.089	0.062	0.144	0.075
	Small	0.048	0.045	0.075	0.058	0.148	0.072
	Large	0.040	0.044	0.065	0.059	0.142	0.072
	Very large	0.035	0.054	0.047	0.065	0.143	0.016

Note *N* = 25

This lack of difference between the intercepts in the shape category indicates that there is no object-level effect for the objects for which the participants had no preexperimental knowledge. A 2 (category) X 4 (intercept) repeated measures ANOVA was performed for the fruit and vegetable categories. There was no significant main effect of category [$F(1,24) = 3.927$, $p = .059$], indicating that there was no difference between the fruit and vegetable categories. There was a significant main effect of intercept [$F(3,72) = 14.359$, $p = .000$]. Contrasts revealed that the bias for relatively very small objects was significantly larger than that for relatively small objects [$F(1,24) = 4.403$, $p = .037$], as was the bias for relatively small versus that for relatively large objects [$F(1,24) = 7.902$, $p = .010$] and the bias for relatively large versus that for relatively very large objects [$F(1,24) = 6.022$, $p = .022$]. These differences in intercepts confirm the prediction that objects studied at the same size are recalled differently depending on the relative size of the studied object. This is consistent with an object-level influence of prior knowledge. The shape category showed no such effect.

We also determined the relationship between recognition response (*remember* or *don't remember*) and recall performance. In this analysis, recall error was assessed by root-mean squared error (RMSE), which measures the absolute deviations between the studied and recalled sizes of the objects by participant.[2] Larger RMSE values indicate worse performance. Figure 5 shows the RMSE by category and recognition response. A 2 (remember) X 3 (category) repeated measures ANOVA was performed. There was a significant main effect of remembering [$F(1,6) = 10.712$, $p = .017$]. Recall error was significantly lower for hits (i.e., *remember* responses) than for misses (*don't remember* responses). There was a significant main effect of category [$F(2,12) = 14.572$, $p = .001$]. Contrasts revealed that recall error was greater for artificial shapes than for

---

[2] NOTE: $RMSE = \sqrt{\sum_i \left(x_i^{recall} - x_i^{study}\right)^2 / N}$, where x^{recall} and x^{study} are the recalled and studied sizes, respectively, and N is the number of objects in the calculation.

fruits [$F(1,6) = 32.155$, $p = .001$] and for vegetables [$F(1,6) = 8.571$, $p = .026$]. There was no significant difference between the recall errors for fruits and for vegetables [$F(1,6) = 4.91$, $p = .069$]. The worst performance was observed for misses in the artificial shape category. In this condition, performance was similar to chance as measured by the mean of the size distribution of the artificial shapes (indicated in Figure 5 by the dotted line). Note that this distribution was by design identical to the vegetable size distribution. There was a significant interaction effect between *remember* responses and category [$F(2,12) = 4.721$, $p = .031$]. Even when participants indicated that they had no conscious recollection of seeing the object at study, they performed better in both the fruit and the vegetable categories than in the artificial shape category.

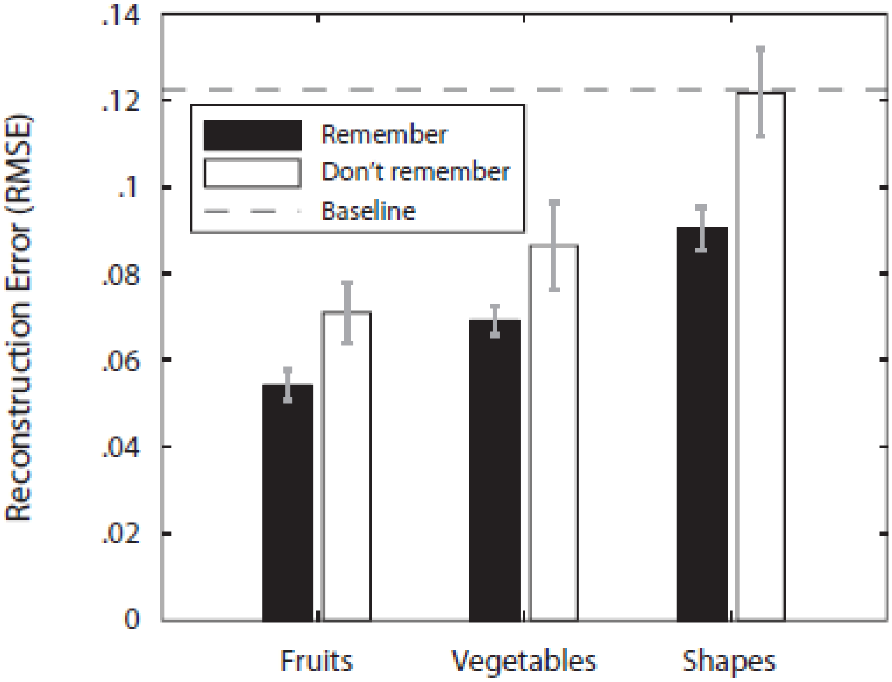


Figure 1.5. Recall error in Experiment 2 measured by root-mean squared error (RMSE). The error is grouped by superordinate category and by recognition response. The baseline error is equivalent to responses based on the mean of the vegetable sizes or, equivalently, the shapes category. Error bars indicate standard errors of the means.

Discussion

We have shown that episodic memories are influenced by prior knowledge and that such influences appear to be hierarchically structured. Objects above the superordinate category mean (e.g., a pineapple) were reconstructed smaller, and objects below the superordinate category mean (e.g., a raspberry) were reconstructed larger. At the same time, an object studied at the small end of that object range (e.g., a small apple) was overestimated, whereas an object studied at the large end of that range (e.g., a large apple) was underestimated, independent of the absolute study size. The biases in size recall are consistent with prior knowledge exerting an influence at both the object-category and the superordinate category levels.

Our experimental design allowed us to rule out various alternative hypotheses about the object-level effects. Sailor and Antoine (2005) hypothesized that temporal contiguity (i.e., sequential effects) produced the center bias effects of the Huttenlocher et al. (1991; Huttenlocher et al., 2000) studies and that reconstruction did not vary systematically as a function of category membership. We, however, have demonstrated that not only does reconstruction vary systematically as a function of category membership but also that these findings cannot be due to sequential effects. Sequential effects cannot be responsible for the object-level effects, because we have demonstrated that objects studied at the exact same size can be recalled differently, depending on the prior knowledge associated with the object. Our findings also preclude the possibility that the object-level bias is due to edge effects and slider start location. These effects would not predict differential bias between vegetables and shapes that are studied at the same size, and they are unable to explain the object-level effects. Therefore, our results show that prior knowledge exerts strong effects on long-term memory and that prior knowledge corrects for noisy episodic memories at the object level. On the other hand, it is not entirely clear whether the negative slopes in Experiment 2 can be solely attributed to prior knowledge at the superordinate-category level. It is possible that additional factors have contributed to this effect. For example, sequential effects can lead to a bias toward the center of the overall size distribution. Similarly, it is not clear whether the participants were utilizing prior knowledge at the superordinate level (e.g., fruits) or at an even more abstract level encompassing all objects encountered during the experiment.

Overall, our results are consistent with the view that the interaction between episodic memory and prior knowledge occurs at multiple levels of abstraction. The nature of the combination of different sources of prior knowledge might depend on the familiarity of the object. Very specific prior knowledge associated with the object might be utilized during recall, when the participant is very familiar with the object. On the other hand, if the object is unfamiliar, prior knowledge at a higher level of abstraction (such as the superordinate level) might be engaged.

In contrast to previous studies, our experimental paradigm allowed us to disentangle recall error for objects that were or were not recognized. One striking finding was that even in the absence of recognition, a smaller reconstruction error was found for objects associated with preexperimental knowledge (fruits and vegetables) than for objects without such knowledge (shapes). This difference cannot be attributed to differences in the size distribution, since the shapes and vegetables were studied at the exact same sizes. In the absence of episodic information about a studied artificial shape, the lowest recall error (as assessed by RMSE) is achieved by responding with the mean of the superordinate category distribution observed in the experiment. The participants appeared to have followed this strategy (see Figure 4). However, in the absence of episodic information about a studied vegetable or fruit, a lower error can be achieved by using both the superordinate and the object size distributions to generate a response. Such knowledge about sizes of specific objects could only have been attained outside the experiment, because in the experimental context, the participants saw a particular object only at a single size. This result supports findings by Huttenlocher et al. (1991; Huttenlocher et al., 2000) that prior knowledge can serve a useful role to improve average performance in recall when episodic memory is weak, incomplete, or error prone.

Using prior knowledge at multiple levels of abstraction is an efficient strategy that allows generalization over experiences and correction of noisy memories. Bartlett's (1932) finding that his British participants tended to recall the natives sailing in a boat rather than a canoe is in line with the predictions of a hierarchical effect that recall for an unfamiliar object, such as a canoe, is influenced by a higher, more abstract level of knowledge— namely, that people sail in boats. In essence, the participants had weak memory traces when recalling the story months or even years after initial exposure but were able to use prior knowledge at a higher level of abstraction to aid recall.

Chapter 2

A Bayesian Account of Reconstructive Memory

Pernille Hemmer & Mark Steyvers (2009). *Topics in Cognitive Science, 1*, 189-202.

Abstract

It is well established that prior knowledge influences reconstruction from memory, but the specific interactions of memory and knowledge are unclear. Extending work by Huttenlocher et al. (1991, 2000) we propose a hierarchical Bayesian model of reconstructive memory in which prior knowledge interacts with episodic memory at multiple levels of abstraction. The combination of prior knowledge and noisy memory representations is dependent on familiarity. We present empirical evidence of the hierarchical influences of prior knowledge, showing that the reconstruction of familiar objects is influenced toward the specific prior for that object, while unfamiliar objects are influenced toward the overall category.

Introduction

Recall of past events can be based on multiple sources of information, such as episodic memories and general knowledge. Information stored in episodic memory provides a direct source of information to guide recall, but this information might be incomplete or noisy. General knowledge about a past event might be helpful for filling in missing details or reducing the effect of noise from episodic memory. For example, when trying to remember the last restaurant we went to and what food we ordered, our recall might be driven not only by vague details of the past restaurant visits but also by general knowledge—perhaps we often go to the same restaurant and we usually order a particular set of dishes. This general knowledge is helpful in guiding recall about past events.

Bartlett (1932) showed that memories are guided by schemas that help to fill in the details of memories. For example, providing additional meaningful information can activate schemas that guide the interpretation of the stimulus and serves as an aid to memory (Bower, Karlin, & Dueck, 1975; Bransford & Johnson, 1972; Carmichael, Hogan, & Walter, 1932). Biases need not be from external sources, but they may arise from internal sources as well (Biernat, 1993; Rehnman & Herlitz, 2006). Bartlett showed that the participants themselves bring certain biases to the task. In both temporal and serial reproduction he demonstrated how a person’s cultural and social experiences influence reconstruction from memory to conform to their idiosyncratic biases. Kalish, Griffiths, and Lewandowsky (2007) formalized Bartlett’s serial reproduction task using iterated learning with Bayesian and human agents. They showed that Bayesian and human learners revert to their prior when inferring the underlying function of a set of coordinates. While serial reproduction is about the evolution from iteration to iteration, the approach presented here will focus on retrieval from memory based on a single specific event.

Previous work by Huttenlocher and colleagues (Crawford, Huttenlocher, & Engebretson, 2000; Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000) has shown that prior knowledge exerts strong influences on reconstruction from memory. Huttenlocher et al. (1991) presented a Bayesian model of category effects positing that reconstruction from memory is a weighted average of specific memory traces and category information. This weighted average ‘‘cleans up’’ noisy memory traces and prevents large errors in reconstruction.

In this paper, we first present the basic approach of the model presented by Huttenlocher and colleagues and then introduce a series of extensions to this model. We assume that the observer is presented with an object during study and is instructed to retrieve from memory a feature of that object at a later time. In the experiment reported in this paper, we test memory for one-dimensional stimulus values, such as the size of an object. In this context, the goal for the observer is to reconstruct the original size μ of an object using noisy samples y that are retrieved from memory. Bayes’ rule gives us a principled way of combining prior knowledge and evidence from memory:

$$p(\mu \mid y) \propto p(y \mid \mu) p(\mu) \qquad (1)$$

The posterior probability $p(\mu \mid y)$ gives the likely stimulus values μ given the noisy memory contents y. This posterior probability is based on a combination of $p(\mu)$, the prior knowledge of the likely sizes of the object and $p(y|\mu)$, the likelihood of obtaining evidence y from memory. This Bayesian approach gives a principled account of how prior knowledge of the world is combined with memory contents to recall information about events.

For example, suppose the feature values of objects are Gaussian distributed, $\mu \sim \mathrm{N}(\mu_o, \sigma_o^2)$, where μ_o and σ_o^2 are the prior mean and variance of the feature values. Furthermore, when a specific object value μ_s is studied, suppose this leads to samples y drawn from episodic memory with the samples having a Gaussian noise distribution centered on the original studied value, $y \sim N(\mu_s, \sigma_m^2)$. The variance of the noise process, σ_m^2, controls the degree to which the stored episodic representations resemble the original studied object features. The exact source of the noise is not modeled in this account, but this could be related to decay or interference with other events entering memory. Standard Bayesian techniques can now be used to calculate the posterior distribution in Eq. 1. The conditional probability of recalled stimulus value μ_r given the contents of memory y is given by a Gaussian distribution with mean μ_n,

$$\mu_n = w\mu_o + (1-w)\bar{y} \tag{2}$$

where $w=(1/\sigma_o^2)/[(1/\sigma_o^2)+(n/\sigma_m^2)]$ and n is the number of samples taken from episodic memory. Note that the mean of the recalled stimulus values is a weighted linear combination of the prior mean μ_o and the mean of memory content $\bar{y}$. The prior mean μ_o is weighted more heavily in recall when the prior has a higher precision $(1/\sigma_o^2)$ and when the memory noise increases. This corresponds to the intuitive notion that if the prior is strong, it will have a strong influence on recall. Similarly, if memory contents are very noisy, the prior will also exert a strong influence on recall.

This model predicts systematic biases toward the category center, or prior category mean, at reconstruction. Figure 1 illustrates these biases and the effect of the strength of the prior. The small vertical lines represent the small noisy samples (around .2) drawn from episodic memory at the time of test. In panels A1–3, we simulate drawing a single sample (n = 1) from memory. The dashed lines represent prior knowledge, and the solid lines represent the posterior distribution that forms the basis for recall. Using classical statistical inference (panel A1) with an uninformative prior, the posterior is centered on the mean of the memory sample—there is no

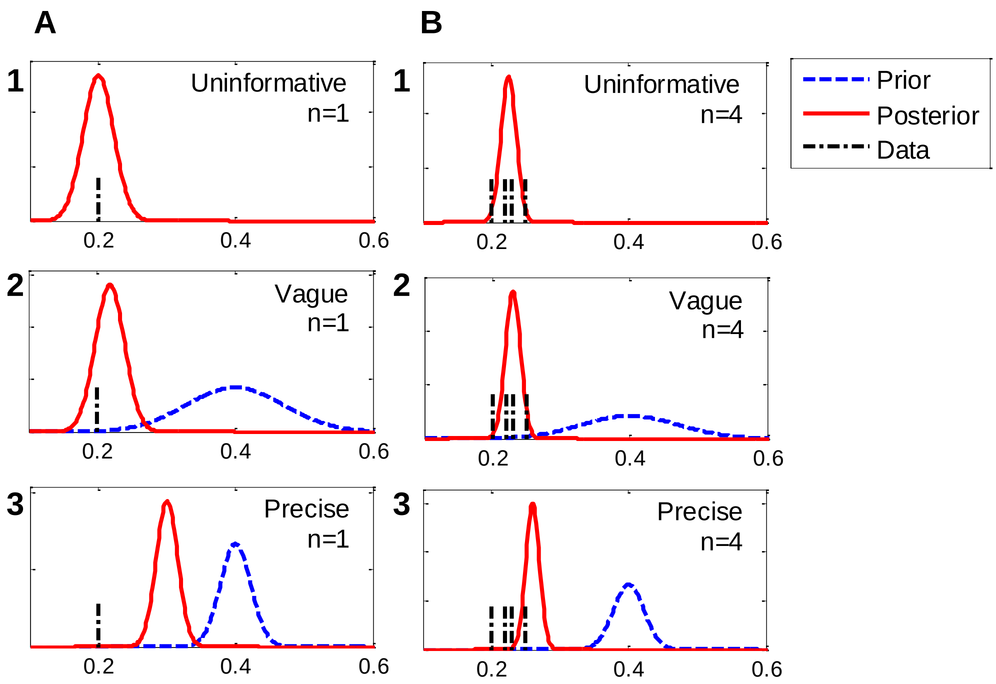


Figure 2.1. Illustrations of a Bayesian account for the systematic biases in reconstructive memory due to prior knowledge. Small vertical lines represent samples drawn from episodic memory. Panels 1A–B show classical statistical inference, panels 2A–B show Bayesian inference with a relatively vague prior, panels 3A–B show Bayesian inference with a relatively precise prior.

effect of the prior. Using Bayesian inference (panels A2 and A3), we specified a prior with mean μ_o = .4. We simulated a relative vague prior with precision $1/\sigma_o^2 = 200$. Using the Bayesian inference procedure as described above, the resulting posterior is slightly shifted toward the prior (A2). For a relative precise prior with precision $1/\sigma_o^2 = 2000$, the result is a posterior that is shifted much more away from the data and toward the prior (A3). Panels B1–3 show the results when four samples (n = 4) are drawn from memory. In this case, the evidence from memory is stronger, which decreases the influence of the prior. Subsequently, the posterior distribution is less influenced by the prior.

Extending the basic approach

The approach sketched above formed the basis for the theory by Huttenlocher and colleagues. We propose an extension to this theory where prior knowledge can come from multiple sources. We will conduct a behavioral experiment using natural objects such as fruits and vegetables for which participants have preexperimental knowledge. We expect that such naturalistic stimuli are associated with more structured knowledge representations where knowledge can be described at several levels of abstraction. For example, we expect that participants not only have prior knowledge at the category level (e.g., ‘‘I expect fruits to be roughly of this size’’) but also at the object level (e.g., ‘‘I expect an apple to be of this size’’). We predict that the influence of the object and category prior knowledge depends on an individual’s familiarity with the object and category. 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 the experiment, we will test some of the predictions from this extended theory and focus on the role of multiple sources of prior knowledge in reconstructive memory.

Experiment

In the following experiment we first measured the perceived size of common natural objects such as fruits and vegetables as an estimate of participants’ prior knowledge for these objects (norming phase). In the second phase of the experiment we assessed recall memory for size. We used the observed size ranges from the norming phase as the foundation for the study sizes in order to encourage the use of prior knowledge.

We predict that the effect of prior knowledge at the category and object level will be observed by systematic deviations toward the mean of the object and category prior at reconstruction. At the category level, this means that small objects (e.g., raspberries) will be overestimated while large objects (e.g., pineapple) will be underestimated. At the object level, this means that objects presented at relatively small sizes (e.g., a small apple relative to all apples) will be overestimated while large objects (e.g., a large apple) will be underestimated.

Separating out the contributions from category and object level is difficult because in many cases, the effects might operate in the same direction. This is illustrated in Figure 2, top panel. If an object is studied at a size that is small relative to both the category and object prior (e.g., a small apple), both of these priors will result in a positive bias. They are both operating in the same direction, toward the category center. The clearest demonstration of independent contributions of object and category level prior knowledge is provided when the effects go in opposite directions. For example, in Figure 2, bottom panel, the object (e.g., a large strawberry) is studied at a size that is large relative to the object prior but small relative to the category prior. In this case, the category effect leads to an overestimation while the object effect will lead to an underestimation of the object at reconstruction. The crucial comparison is between the top and bottom panel. In both cases, the objects are shown at exactly the same size during study. Because both the category and object level effects might operate simultaneously, it is unclear what the combined result is. However, we predict that objects that are studied at the same absolute study size can be differentially biased when the objects are associated with different object-level prior knowledge.

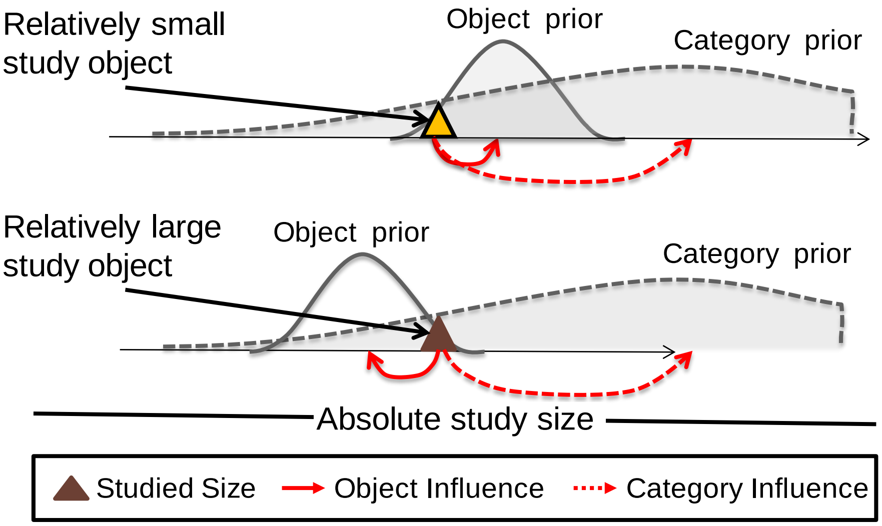


Figure 2.2. Predicted influences of category and object level priors for two objects studied at the same size.

Methods

Participants

Participants were undergraduate students at the University of California, Irvine. There were 18 participants in the norming phase and 25 participants in the test phase.

Materials and procedures

For the norming phase there were 37 images in two categories: fruits and vegetables. For the test phase, 24 of the objects from each of the norming categories were used. See Figure 4 for examples. Another class of stimuli was also developed: abstract shapes created by drawing outlines of objects and filling with blue. See Figure 3 for examples.

Norming phase

All materials were presented on two computer screens. A reference object was presented on the left screen and the object of interest was presented on the right screen. Participants were

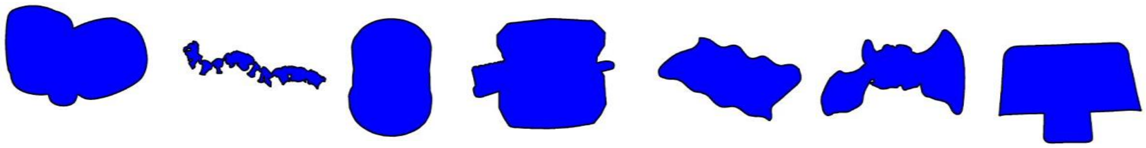


Figure 2.3. Examples from the shapes category created by drawing outlines of objects filled in blue.

asked to make three size judgments for each object: ‘‘What is the smallest (or average or largest) size of an object like this?’’ Participants manipulated the size of the object using a slider. Responses were measured on a scale from 0 to 1 (where 1 corresponds to the maximum size of the computer screen). Images were blocked by category (fruits and vegetables) and presented in random order within each block. Images were initialized at 1 of 4 sizes relative to the overall screen size: .2, .4, .6, or .8.

Memory phase

The study sizes of the images were sampled from the size ranges collected in the norming phase. For sampling, we used a truncated Gaussian distribution between the minimum and the maximum of the individual object range. The objects were never shown outside of the minimum–maximum range. The shapes category was yoked to the vegetable category for size and orientation on the screen. The specific study size for each shape was the same as that of its yoked vegetable. Participants were shown a continuous random sequence of study and test images. The images were blocked by category (fruits, vegetables, and shapes). Each study image within each category was presented a total of three times during the experiment, and there was always a related intervening test trial between presentations. The study size for a given image remained constant across the three repetitions. Each participant completed three blocks of 72 study and test images, and trials and blocks were randomized across subjects. Study images were presented for 2 s. At test participants were asked to make two memory judgments. They were first asked to make a recognition decision about whether they remembered seeing the object at study. Second, they were asked to make a recall judgment about the size of the object at study using the slider on the screen to manipulate the size of the object. The object remained on the screen until the participant was satisfied with the current size judgment and clicked a button to continue. Responses were measured on a scale from 0 to 1. On test trials, the images were shown randomly at 1 of 4 sizes: .2, .4, .6, and .8. The slider was initiated at the corresponding location.

Result

Norming phase

Figure 4 depicts the 24 objects from the vegetable category. The top panel indicates the range of the size judgments for individual objects averaged over participants. The results follow a natural order: the mean ‘‘average’’ size judgment for mushrooms is smaller than for bell peppers, which are all smaller than celery, and so on. Participants expressed a large degree of agreement, although variability does increase with the magnitude of the objects.

Memory phase

Reconstruction error (reconstructed size—studied size) was used to measure performance in each category. Reconstruction error as a function of category and object class is plotted in Figure 5. Positive reconstruction error indicates overestimation while negative reconstruction error indicates underestimation. The observed pattern of correction toward the category center as indicated by negative slopes for all categories supports the prediction of category effects.

To assess the influence of object priors, we divided the study objects into four classes based on the study sizes relative to the minimum and maximum acceptable sizes as assessed in the norming experiment. We divided the range between the minimum and maximum in four equal ranges and named those ranges ‘‘very small,’’ ‘‘small,’’ ‘‘large,’’ and ‘‘very large.’’

Figure 6 illustrates this discretization process. These classes therefore give the sizes of objects relative to the mean of the object—for example, a ‘‘very large’’ object might be an apple that is studied at close to the maximum size (relative to all apples). For each of the four relative object classes we divide the absolute study sizes into the top, middle, and bottom 33% percentile. Mean reconstruction error was calculated separately for the trials that fell into these top, middle, and bottom bins. Because we applied this procedure separately for each relative object size, the absolute study sizes do not line up between relative object sizes (see Figure 5).

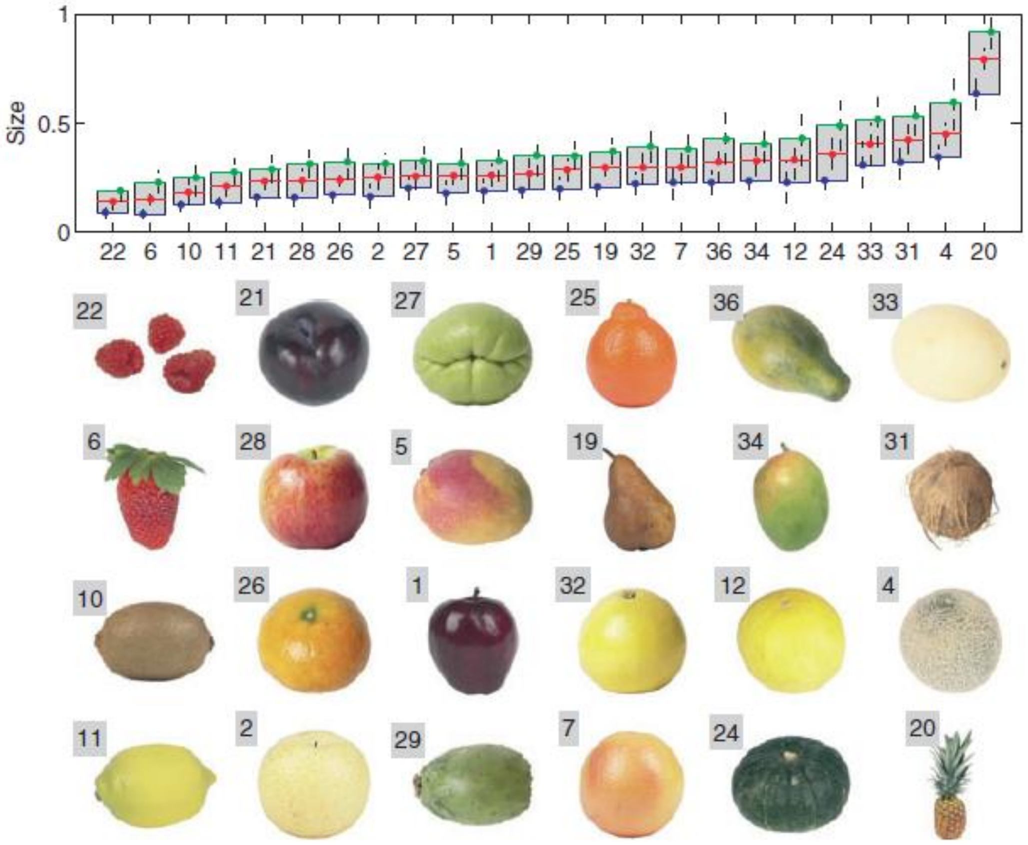


Figure 2.4. Norming results for the fruit category. The objects are in reading order by the mean of the ‘‘average’’ rating. The bar graph shows the range of size judgments for individual objects. The scale on the vertical axis ranges from 0 to 1, where 1 corresponds to the maximum size of the computer screen. For each bar, the mean of the ‘‘small,’’ ‘‘average,’’ and ‘‘large’’ sizes are indicated by the bottom line, middle line, and top line, respectively. The vertical lines correspond to the 25–75% confidence interval across participants.

The results show a regular pattern for different object classes (very small < small < large < very large). This difference in intercepts by relative study size supports the prediction of object prior effects. To measure the effects of prior knowledge on reconstruction memory at both the category and object level, a regression model was fitted to each subject assuming a fixed slope and separate intercepts for each relative object size. Average slopes and intercepts are reported in Table 1.

The slope for each category was significantly different from zero (fruits: t[24] = -4.714, p = .000, vegetables: t[24] = -5.657, p = .000, shapes: t[24] = -10.754, p = .000). This is consistent with a category-level influence of prior knowledge. A repeated measures anova for the fruit and vegetable categories showed a significant main effect of intercept (F[3.72] = 14.359, p = .000). A significant trend was observed such that intercepts for very small objects

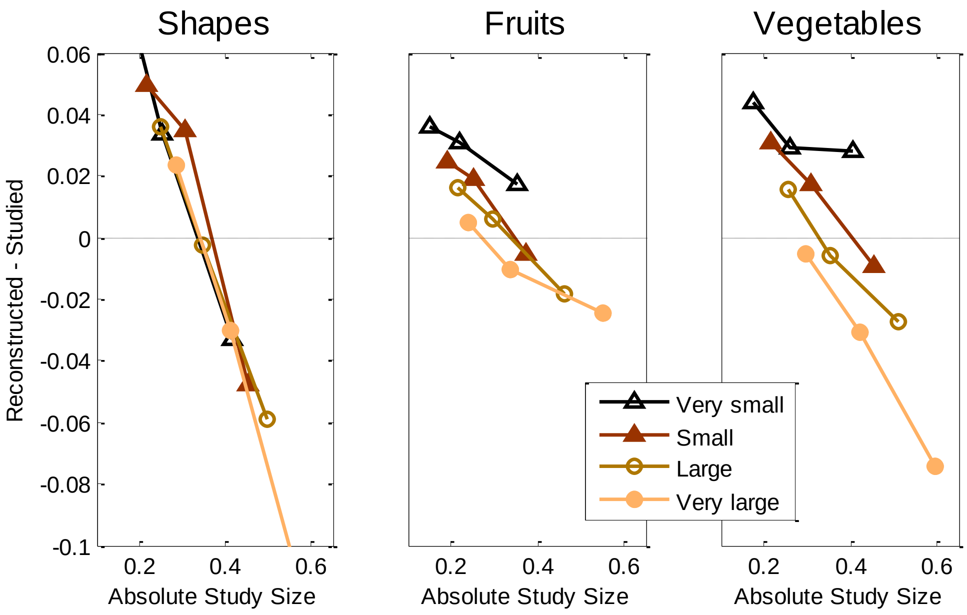


Figure 2.5. Reconstruction error as a function of category and object class. Study objects are divided into four classes relative to the mean of the object: ‘‘very small,’’ ‘‘small,’’ ‘‘large,’’ and ‘‘very large.’’ Negative slopes show correction toward the category center and are indicative of category-level knowledge effects. Intercept differences by relative study size are indicative of object-level knowledge effects.

Table 2.1. Average slopes and intercepts by category.

	Fruits			Vegetables			Shapes		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Slope	-.120	.127	25	-.191	.169	25	-.415	.193	25
Intercepts							.142	.069	25
Very small	.057	.049	25	.089	.062	25			
Small	.048	.045	25	.075	.058	25			
Large	.040	.044	25	.065	.059	25			
Very large	.035	.054	25	.047	.065	25			

were larger than that of relatively small objects (F[1,24] = 4.403, p = .037), as was that of relatively small to relatively large objects (F[1,24] = 7.902, p = .010), and relatively large to relatively very large objects (F[1,24] = 6.022, p = .022). These differences are consistent with an object-level influence of prior knowledge. A repeated measures anova for the shape category found no effect of intercept (F[3,72] = .453, p = .716). This lack of difference between the intercepts in the shape category indicates that there is no object-level effect for the objects for which the participants have no preexperimental knowledge.

Model

The results showed that natural stimuli such as fruits and vegetables are associated with multiple levels of preexperimental prior knowledge, each exerting an influence on reconstructive memory such that objects studied at the same absolute size can be differentially biased depending on the prior knowledge at the object level. In our first extension of the basic model by Huttenlocher and colleagues, we propose that prior knowledge can be represented at multiple levels of abstraction which can independently influence reconstruction from memory. We propose a simple mixture model where the prior mean and variance (μ_o, σ_o^2) is a combination of category and object level priors,

$$\mu_o = z\mu_i + (1 - z)\mu_c \tag{3}$$

$$\sigma_o^2 = z\sigma_i^2 + (1 - z)\sigma_c^2 \tag{4}$$

where (μ_i , σ_i^2) represents the object prior associated with object i and (μ_c , σ_c^2) represents the category prior. The variable z weights the contribution of the object prior relative to the category prior. We assume that this weighting is determined by

$$z \sim \text{Bernoulli}(\theta_i) \tag{5}$$

where θ_i is a constant that represents the familiarity of an object. In this model, familiar objects lead to a prior that is more dependent on the object rather than the category. Similarly, this implements the intuitive notion that for unfamiliar objects, it is unlikely that the object prior is reliable and inference instead reverts to a higher-level prior based on categorical knowledge.

As before, we assume that the computational goal for the participant is to invert the forward memory model and reconstruct the original event given the noisy memory contents and prior knowledge about the study event. The solution to this computational problem was described in Eq. 2.

We applied the model in Eqs. 2–5 to our experimental setup and aimed for qualitative fits to the data as opposed to detailed quantitative fits. We used three values of θ corresponding to no object familiarity (θ = 0), medium familiarity (θ = .4), and high familiarity (θ = .7). In this model, the category prior can represent a combination of a priori knowledge about the category as well as knowledge accumulated during the experiment (such as the distribution learned for the

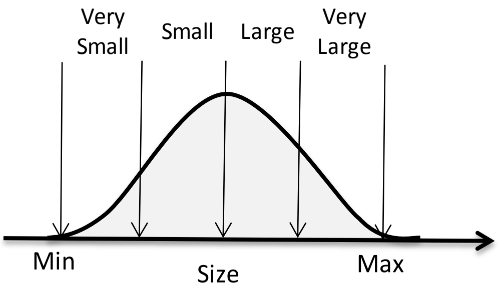


Figure 2.6. Example of the Gaussian distribution for a simulated object prior. The four regions label the study sizes relative to the object prior.

shapes category). Here, we will not distinguish between these two sources for the category prior and use a single prior with $\mu_c = .5$ and $1/\sigma_c^2 = 20$. This is a relatively vague prior that is centered near the mean of study sizes we used in the experiment. For the object priors, we simulated Gaussians with means centered across the range [0,1] and precision $1/\sigma_i^2 = 200$. This implements a relatively precise object prior compared to the category prior. For the study sizes, we drew samples from the object priors and rejected samples outside the [0,1] range. For the purpose of data analysis, we categorized the study sizes into four classes: ‘‘very small,’’ ‘‘small,’’ ‘‘large,’’ and ‘‘very large’’ (as illustrated in Figure 6). These size indications are relative to the object prior. Just as in the experiment, the label ‘‘very small’’ refers to an object that was presented at study at a value close to the minimum size for that particular object. This size is not related to the absolute study size. For example, we can simulate a very small pineapple that is still larger than most other fruits. Finally, we ran the simulation with a memory precision of $1/\sigma_m^2 = 50$.

Figure 7A shows the model predictions. Overall, the results show effects of both the category and object prior. Objects that were studied at small sizes with respect to the category and the object prior are overestimated while large study sizes relative to the object and category prior are underestimated. Also, as expected, variations in familiarity can modulate the influence of the object prior. For h = 0, there is no influence of the prior. This situation is comparable to the experimental results for the shapes category for which participants did not have any preexperimental knowledge specific to the object.

When comparing the slopes in this simulation and the experimental results in Figure 5, an important discrepancy arises. In the experimental data, the effect of the category prior is stronger for the shapes compared to the fruits and vegetables (see Table 1 for the difference in estimated slopes across categories). In the simulation in Figure 7A, the effect of the category prior is more or less constant across the levels of familiarity. The difference in category prior cannot be explained due to differences in the study size distributions because the shapes category was yoked to the vegetables category and exactly the same sizes were presented across the two categories.

These experimental results raise an interesting issue about the relative effects of the priors. For objects that have presumably very little prior knowledge, we see relative strong effects of the priors, exactly opposite to what the basic Bayesian approach would predict. This suggests that an additional change to the theory is needed to fully explain the data. We will now describe a change to the noise process that governs the sampling of memory representations. This additional extension will lead to a model that is able to qualitatively describe our findings.

In the basic approach, the memory noise $\sigma_{\mathrm{m}}{}^2$ is treated as a constant parameter and the theory does not explain how this parameter is set or varies across experimental conditions. Moreover, this approach assumes that the observer knows the memory noise parameter during the inference process. However, it seems unlikely that the observer has access to such knowledge. We propose that the memory noise is itself an unobserved variable which needs to be estimated from the data (i.e., the memory samples) and prior knowledge. From a statistical point of view, we propose a system where the goal is to make inferences about data with an unknown

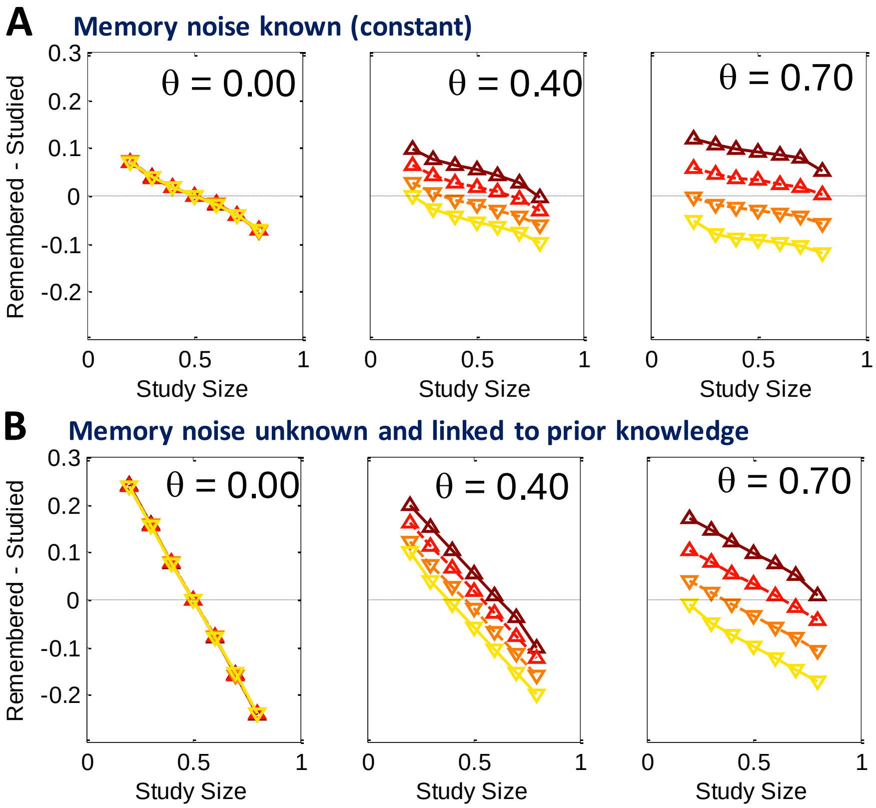


Figure 2.7: (A) model predictions when memory noise is a constant parameter. (B) model predictions when memory noise is an unknown variable.

mean and unknown variance. In the statistics literature, several solutions exist for this problem, and we follow a standard approach (e.g., Gelman, Carlin, Stern, & Rubin, 2003) that allows an analytic solution.

As before, we assume that noisy memory samples y are drawn from episodic memory with a Gaussian noise distribution $y \sim \mathrm{N}(\mu_s, \sigma_\mathrm{m}^2)$ that is centered around the original studied value μ_s. Instead of assuming a constant noise variance σ_m^2, the noise variance is sampled from an inverse-χ^2 distribution:

$$\sigma_m^2 \sim \text{Inv-}\chi^2\left(\nu_0, \sigma_0^2\right) \tag{6}$$

and the mean of the stimulus values is assumed to be conditionally dependent on the noise variance:

$$\mu_s \mid \sigma_m^2 \sim N\left(\mu_0, \sigma_m^2 / \kappa_0\right) \tag{7}$$

The constants υ_0 and κ_0 represent the prior degrees of freedom and the prior sample size, respectively. The goal for the observer is to calculate the conditional probability of recalling size μ_s given the contents y in memory. For notational purposes we will refer to this conditional probability as μ_r, where $\mu_r \sim \mu_s \mid y$. This leads to the following solution:

$$\mu_r \mid y \sim t_{v_o+n}(\mu_n, \sigma_n^2) \tag{8}$$

$$\mu_n = \frac{k_o}{k_o+n}\mu_o + \frac{n}{k_o+n}\bar{y} \tag{9}$$

$$\sigma_n^2 = \frac{v_o\sigma_o^2 + (n-1)s^2 + \left(k_0 n / k_o + n\right)(\bar{y}-\mu_o)^2}{(v_o+n)(k_o+n)} \tag{10}$$

Note the similarity of Eq. 9 to Eq. 2. In both cases, the mean of the recall distribution is a linear combination of the prior mean and the mean of the observed memory samples. Figure 8 shows a graphical representation of the complete model. Shaded nodes represent observed variables while nodes without shading represent unobserved variables. The arrows indicate the conditional dependencies between the variables.

Note also that memory noise is modeled as an unobserved variable and that there is a coupling between the memory noise variance and the prior variance of the study event. This corresponds to intuitive notions about memory; if we encode objects with which we are not very familiar and have little associated prior knowledge, it is more difficult to store accurate representations in memory for that object. In contrast to the previous model where memory noise was left as an unexplained parameter, this model explains memory noise as a variable dependent on prior

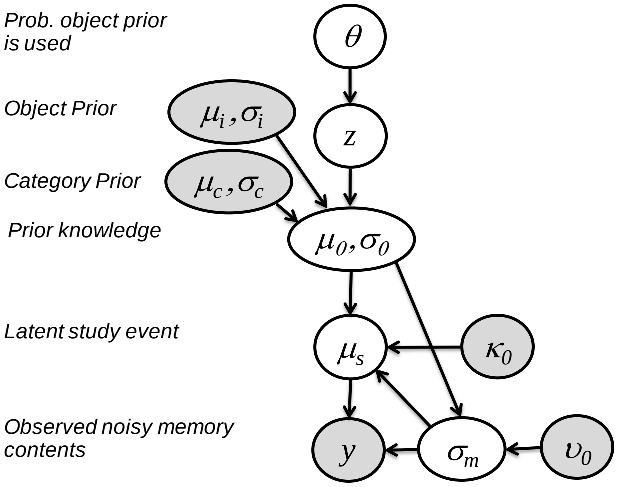


Figure 2.8. The graphical model representation for the hierarchical Bayesian model for reconstructive memory.

knowledge. We simulated this model in the same manner as the previous model. We used the same category and object priors and set $\upsilon_0 = 5$ and $\kappa_0 = 5$. The results are shown in Figure 7B. Note that the model predicts that the category prior is relatively strong for the low familiarity conditions. This somewhat paradoxical effect falls out of the model because of the coupling between memory and noise and prior knowledge. Objects with weak priors (e.g., shapes) are associated with relatively noisy samples from memory. The result is that the prior exerts a stronger influence to reduce the effects of the memory noise. On the other hand, objects with strong priors (e.g., fruits and vegetables) are associated with relative precise samples from memory, leading to a reduced influence of the prior overall.

Conclusion

We have given a Bayesian account of reconstructive memory, where reconstruction of the size of the original study event is influenced by prior knowledge at multiple levels. This follows from our empirical evidence that reconstruction from memory is influenced by prior knowledge at multiple levels of abstraction. Unfamiliar objects lead to inferences that are more influenced by the category center, whereas familiar objects lead to inferences that are more influenced by the object prior. A novel assumption of the model is that memory noise is unknown to the observer and becomes part of the inference process. This assumption is different from the basic approach as described by Huttenlocher et al. (1991, 2000) but is consistent with empirical data showing that category effects exert a greater influence when the observer has no preexperimental knowledge for the object, that is, the object is unfamiliar. While it seems counterintuitive that a vague prior exerts a stronger influence in reconstructive memory, this is to be expected if we couple the memory noise process to the prior.

Chapter 3

Integrating Episodic and Semantic Information in Memory for Natural Scenes

Pernille Hemmer & Mark Steyvers (2009). *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Abstract

Recall of objects in natural scenes can be influenced not only by episodic but also by semantic memory. To model the statistical regularities that might be encoded in semantic memory, we applied a topic model to a large database of labeled images. We then incorporated the learned topics in a dual route topic model for recall that explains how and why episodic memories are combined with semantic memories. The dual route model was applied to an empirical study in which people recall objects from scenes under varying amounts of study time. The dual route model explains how the trade-off between episodic and semantic memory is affected by study time, output position, and also congruity of the object with the scene context.

Introduction

Semantic knowledge can exert strong influences on episodic recall. In the verbal domain, the use of highly related words on a study list can lead to intrusions of related 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. For example, people can recall seeing books in an office where there were no books present (Brewer & Treyens, 1981). These intrusions demonstrate the influence of semantic knowledge on recall. Some researchers have viewed such intrusions as demonstrations of shortcomings of the memory system. However, semantic knowledge can also serve as an aid to episodic memory and lead to improvements in recall performance (e.g. Hemmer & Steyvers, 2009; Konkle & Oliva, 2007; Huttenlocher et al. 1991).

Dual retrieval accounts of memory propose that reconstruction from memory requires accessing either the verbatim memory trace or semantic information relevant to the event (Brainerd et al., 2002). The verbatim – or episodic memory – trace is a representation close to the original event, while the semantic information is an abstraction of the event, often referred to as ‘gist’ or ‘schema’. Previous dual route models have not explained in detail how the semantic information is represented (or extracted from the environment) and have not fully described the detailed mechanisms for the interaction between episodic and semantic information.

In this research, we build on the framework of rational memory models that assume that the memory system is exploiting environmental regularities when recalling information about past events (Anderson, 1990; Steyvers & Griffiths, 2008). We develop a dual route memory model and apply it to the problem of recalling objects from natural scenes. We assume that an observer is presented with a scene during study and is instructed to retrieve from memory objects that occurred in the scene. The goal for the observer is to reconstruct the objects from the scene optimally combining the available information. We assume that the available information is based on noisy episodic memories and also on encoding based on the semantic context. Previous research has shown that people are sensitive to the contextual information in scenes and can quickly extract a high-level semantic representation of a scene (Potter et al., 2002).

In this paper, we will first present an empirical study on scene recall and investigate how recall accuracy varies as a function of study time and what the accuracy is if there is no episodic information at all and recall is based on semantic information only. The experimental data allow us to assess how people trade off between episodic and semantic memory. We then present a topic modeling analysis (Griffiths & Steyvers, 2004; Griffiths, Steyvers & Tenenbaum, 2007) for a large database of labeled images. The extracted topics serve as approximations to the kinds of statistical regularities that people might have encoded in semantic memory. Lastly, we will show how a dual route topic model (Steyvers & Griffiths, 2008) that mixes episodic and semantic information during encoding can account for the empirical findings. We also show how the model can explain the Von Restorff effect, where people have better memory for objects that are incongruous with the scene context.

Empirical study on scene recall

We conducted a series of behavioral experiments using natural scenes such as kitchens and offices to quantify the relative contribution of semantic knowledge on recall. In a memory experiment, we showed images of natural scenes for varying amount of study time. We expected that by decreasing the amount of study time, recall would be based more on semantic memory and would lead to a larger number of errors. To assess the prior knowledge people have about certain types of scenes, we also conducted a norming study where we asked participants to name the objects they expected to appear in certain types of natural scenes, without actually showing them any image. Finally, we ran a perception experiment, using the same images as used in the memory experiment, where participants were asked to name all the objects that they perceived in the image. This perception experiment allowed us to assess the ground truth of which objects were perceived to be present in each image, which can be used to score the accuracy of responses in the memory experiment.

Methods

Participants were undergraduate students at the University of California, Irvine. There were 22 participants in the prior knowledge experiment, 25 participants in the perception experiment, and 49 participants in the memory experiment.

Materials

We sampled 10 images from the LabelMe database (Russel & Torralba, 2008) where we chose 2 images each of 5 different scene types. The scene types correspond to kitchen, dining, office, hotel room, and urban scenes.

Prior Knowledge Experiment

To assess prior (semantic) knowledge about specific scenes, we asked participants to list objects that they would expect to occur in a given scene type (which was described by the verbal label). Participants entered their responses on a computer screen and were required to make responses for a minimum of 60 seconds before continuing to the next question.

Perception Experiment

In this experiment, we assessed the ground truth for the occurrence of objects in each of the 10 images. Materials were presented on two computer screens. The image was presented on the left screen while response instructions and a response box were presented on the right screen. Participants were asked to list the objects present in each image and were required to make responses for a minimum of 60 seconds. They received feedback based on matching their responses to those of previous participants. Images were presented in random order.

Memory Experiment

For the memory experiment, participants studied an image for either 2 or 10 seconds. After completing a short distracter task, participants were asked to list all the objects they recalled seeing in the presented image. Study images were presented in random order. 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.

Response Normalization

Responses for all experiments were corrected for spelling, plurals, 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 mapped as "car".

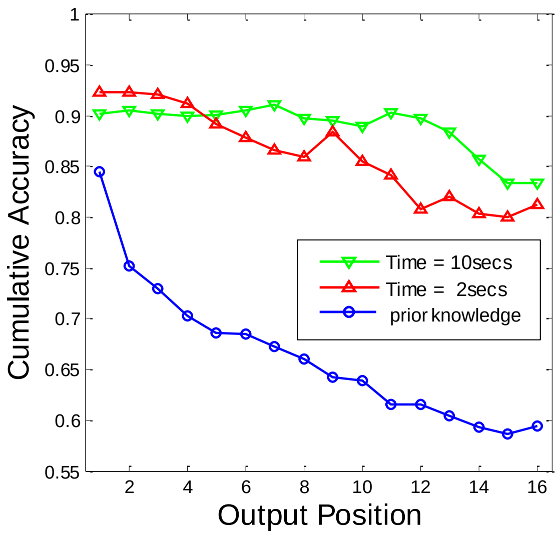


Figure 3.1. Cumulative accuracy as a function of study time and output position. The figure also shows the simulated performance when one treats the responses from the prior knowledge experiment as responses in the memory experiment.

Results and Discussion

To measure performance in the memory experiment, we checked whether a given recalled object was part of any of the responses that were given by participants in the perception experiment. If it was, it was scored as a correct response. If it was not, we manually checked whether the recalled object could still be considered as a description of an object that was part of the image. Only if it was not, the response was scored as incorrect. We calculated cumulative accuracy in the memory experiment as a function of the output position. In other words, we calculated the mean accuracy for the first item recalled, first two items recalled, etc. Figure 1 shows the cumulative accuracy as a function of output position and study time. Overall, cumulative accuracy decreases as a function of output position. Therefore, more intrusions are made 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 second 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.

Figure 1 also shows the performance one can expect from prior knowledge in the absence of any episodic information. This is the case where the image was not studied at all (corresponding to zero second study time). Even though we did not actually run this in the memory experiment, we can consider the responses from the prior knowledge experiment as reasonable guesses to the objects of an image in a particular scene. We ran an analysis 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. Figure 1 shows that the performance of 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 associated with any episodic knowledge of the task. 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, which might be smaller than one might expect. These results demonstrate that general knowledge of scenes can greatly contribute to the accuracy of recalling objects from natural scenes.

A Model for Object Recall in Natural Scenes

One conclusion from our empirical study is that semantic knowledge can lead to good baseline performance in scene memory. 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. This raises the question of how the interaction between episodic and semantic memory can be modeled. We will first discuss a topic model for scenes that approximates the semantic knowledge people might have about objects in scenes and then develop a dual route topic model that integrates both episodic as well as semantic memory information.

A Topic Model for Scenes

Probabilistic topic models have been developed as a method to automatically learn semantic representations for documents by analyzing the statistical relationships between words and the documents they occur in (e.g. Griffiths & Steyvers, 2004; Griffiths, Steyvers & Tenenbaum, 2007). In the topic model, each document is expressed as a mixture of topics that can be thought of as the gist of a document, and each topic represents a probability distribution over words. Here, we apply the topic model to a subset of 13,572 images of the LabelMe database (Russel & Torralba, 2008). These images were annotated by volunteers resulting in a total of 87,152 labels and 3782 unique types. The subset contains images of natural scenes, such as urban street scenes and indoor scenes of kitchens and offices.

We treat each scene from the database as a mixture of topics and each topic as a distribution over image objects. This specifies a generative model in which objects in a scene are selected by first sampling a topic from the topic distribution associated with the scene and then sampling an object from the topic. Specifically, the conditional distribution of an object o in a scene s is given by,

$$P(o \mid s) = \sum_{t=1}^{T} P(o \mid z=t) P(z=t \mid s) \qquad (1)$$

where $p(o|\ z=t)$ is the multinomial distribution over objects given topic t and indicates which objects are important to a topic, and $p(z=t|\ s)$ is the multinomial distribution over topics given scene s and indicates which topics are important to a particular scene.

Figure 2, panel A shows a graphical model representation of the topic model. Shaded nodes represent observed variables while nodes without shading represent unobserved variables. The arrows indicate the conditional dependencies between the variables, and the plates show the replications of sampling steps. There are S scenes and each scene has N_s objects. The variable θ is the scene-topic multinomial and ϕ is the topic-object multinomial. The priors on the multinomials are Dirichlet distributed with hyperparameters α and β. We treat α and β as constants in the model (we set $\alpha = 0.1$ and $\beta = 0.01$).

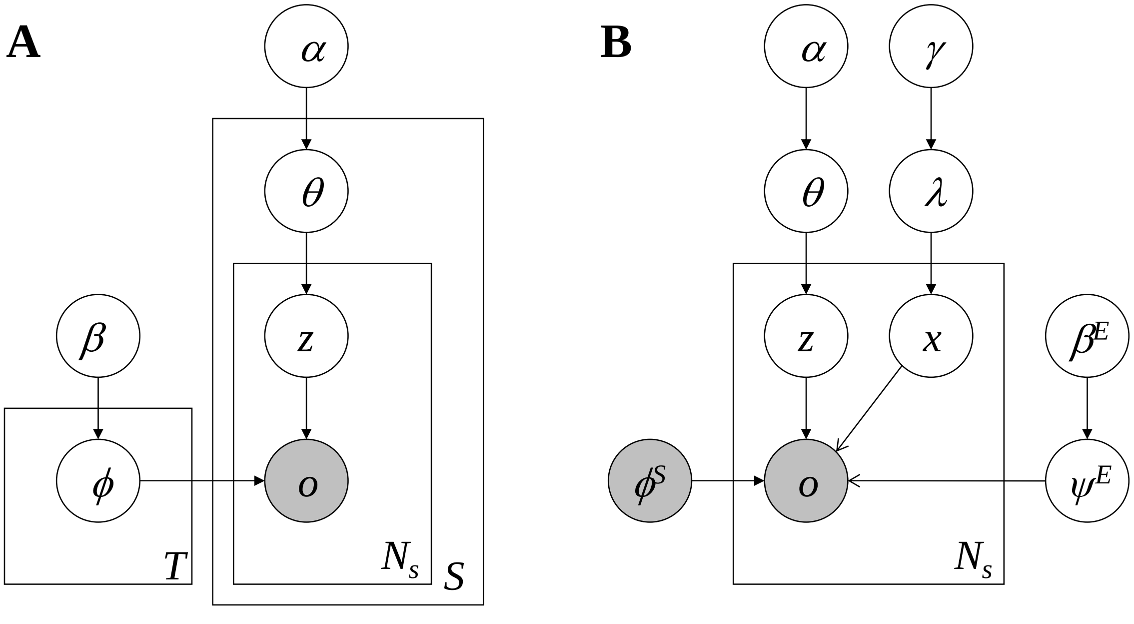
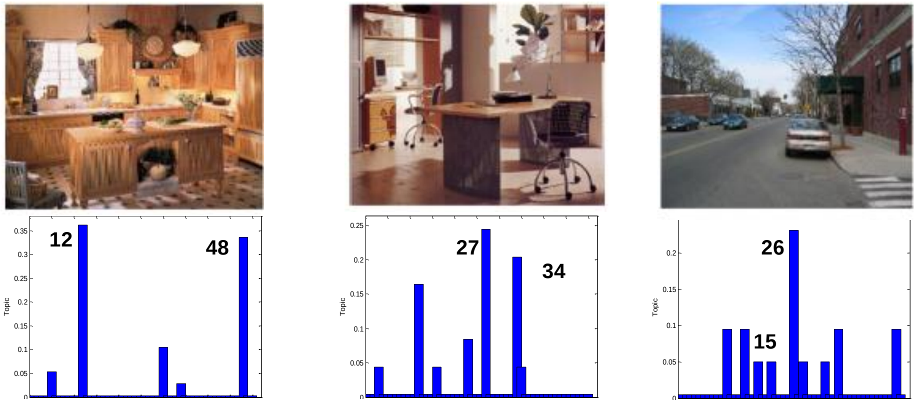


Figure 3.2. The graphical model representation for A) the standard topic model and B) the dual route topic model.



Topic 12	Topic 48	Topic 27	Topic 34	Topic 26	Topic 15
cabinet	***chair***	screen	***lamp***	person	***tree***
stove	box	***keyboard***	***wall***	***building***	grass
table	book	mouse	bed	***sign***	***sky***
sink	***plant***	***table***	***table***	***car***	bench
fridge	***picture***	mug	***floor***	streetlight	***plant***
wall	lamp	mousepad	phone	***window***	person
floor	***basket***	***cpu***	curtain	***road***	path
faucet	bookshelf	speaker	pillow	***sky***	***building***
dishwasher	painting	phone	***chair***	flag	curb
coffeemachine	***cabinet***	***chair***	painting	***sidewalk***	streetlight

Figure 3.3. Model predictions for three scene types: kitchen, office and urban. The bar graphs show the distribution over 50 topics for a scene with topic indices for the two most likely topics. The rank-ordered object distributions corresponding to these topics are shown below. Objects labeled in bold were part of the original image annotations.

We applied the topic model with T=50 topics to the LabelMe image database and used Gibbs sampling to infer both $p(o|\ z=t)$ and $p(z=t|\ s)$. Several examples of topic distributions are illustrated in figure 3. The figure shows images from three different scene types: kitchen, office, and urban with the inferred topic distribution for that image. For example, topic 12 is the most likely topic for the particular kitchen image and topic 27 is the most likely topic for the particular office image. Some of the likely topics are illustrated in at the bottom of figure 3. This shows the list of most likely objects associated with each topic. Overall, the model shows that the topics for each image qualitatively capture the semantic context of the image. The likely objects in the topics associated with scenes are objects that can reasonably be found in the respective scenes, and seem to describe the 'gist' of the scene.

A Dual Route Model for Object Recall

The topic model itself cannot be a complete model for reconstructive memory. The topic distribution for a scene provides a generalized representation for the occurrence of objects in scenes (e.g., offices), which is useful to characterize the "gist" of a scene. However, the distribution over topics is insufficient to represent the exact set of objects present in an image. In human memory, recall can be quite accurate, if given enough study time. Therefore, to give a more complete account of human memory, we need to expand the topic model with an additional component that allows the model to reconstruct the specific objects present in a scene.

We will now describe an extension to the standard topic model called the dual route topic model introduced by Steyvers and Griffith (2008). We will apply the model to the problem of scene recall. In the model, recall of objects in a scene is a result of two processes: episodic recall and recall based on the semantic context. The semantic information is an abstraction based on the statistical regularities of the collection of scenes. For each image, the semantic context is encoded by a probability distribution over topics. The episodic information is based on a noisy encoding of the actual list of objects present in the image to be remembered. In the model, we will implement episodic noise by a simple sampling process. We assume that the episodic sampling process is based on a multinomial distribution over objects with a symmetric Dirichlet prior,

$$o|\psi \sim \mathrm{Mult}(\psi) \qquad (2)$$

$$\psi \sim \mathrm{Dirichlet}(\beta^E) \qquad (3)$$

where β^E is the hyperparameter that controls the amount of smoothing. Note that this process is not just defined over the observed objects in the image, but over all object types (i.e., the object vocabulary). In this process, it is possible to give high probability to a variety of objects, making them likely to be retrieved from the episodic route. However, with the Dirichlet prior, a capacity constraint can be built in. With small values of β^E, it is unlikely, a priori, that the probability over objects is distributed over a large number of objects, therefore encouraging a sparse representation of objects. Therefore, the smoothing parameter determines how much of the retrieval process focuses on the observed objects versus other objects in the vocabulary.

If recall is based strictly on this episodic component, performance should be accurate, at least for a subset of items on the list, but it could potentially fail to fully retrieve the whole list. If recall is based strictly on semantic information performance might not be as accurate but the topic distribution allows retrieval of a larger number of items. The dual route topic model allows recall to be a mixture of these two extremes. The weighting is such that recall is neither too specific nor too general. In the model a mixing process determines if an object is generated using the episodic route or using semantic information. An indicator variable *x*, acts as a switch such that if *x*=1, the object is sampled from the semantic route, and if *x*=0, the object is sampled from the episodic route. We assume that the probability of a route assignment is distributed Bernoulli with a symmetric Beta prior:

$$x \mid \lambda \sim \text{Bernoulli}(\lambda) \tag{4}$$

$$\lambda \sim \text{Beta}(\gamma) \tag{5}$$

Therefore, the conditional distribution of an object *o* given a scene *s*, is given by:

$$p(o \mid s) = p(x = 1 \mid s) \sum_{t=1}^{T} p(o \mid z = t) p(z = t \mid s) + p(x = 0 \mid s) p'(o \mid s) \tag{6}$$

where the first term is the distribution over objects predicted by the topic model weighted by the probability of a route assignment in favor of a semantic encoding. The second term is the object distribution *p*'(*o*| *s*), predicted by the episodic route weighted by the probability of a route assignment in favor of an episodic encoding.

Note that this model specifies a generative procedure for producing objects in a given scene. Figure 2B shows a graphical representation of the complete model. Note that we assume that the distribution over objects in each topic, ϕ^S, is observed and estimated by the topic model in a prior learning phase.

The main use of the model is as an encoding model where the goal is to infer the encoding parameters conditional on the observed set of objects in an image. In other words, the goal is to find an encoding such that during retrieval, the model is likely to reconstruct the observed set of objects in an image, taking into account the probabilistic constraints of the model – the built in capacity constraint for the episodic route and the overgeneralization of the semantic route. Because the model assumes that each object originates from a single memory route, the goal of encoding is to infer which objects can be encoded via the episodic route and which objects can be reconstructed by a probability distribution over topics (specific for the image studied).

The latent variables *z* and *x* can be inferred using Gibbs sampling (the remaining latent variables can be integrated out). The topic and route assignment for the *i*-th object can be jointly determined conditional on all other assignments:

$$p\left(x_i = r, z_i = j \mid z_{-i}, x_{-i}, \mathbf{o}, \varphi^S\right) \propto \begin{cases} \phi^S_{j,o_i}\left(n_{j,-i}+\alpha\right)\left(n_{1,-i}+\gamma\right) & r = 1 \\ \dfrac{n_{o_i,-i}+\beta^E}{n^V_{-i}+M\beta^E}\left(n_{j,-i}+\alpha\right)\left(n_{0,-i}+\gamma\right) & r = 0 \end{cases} \tag{7}$$

where M is the number of unique objects labels in the LabelMe database, $n_{0,\text{-i}}$ is the number of time the episodic route is assigned, $n_{1,-i}$ is the number of times the semantic route is assigned, and $n_{o,-i}$ is the number of times a specific object o is assigned to the episodic route. The subscript $-i$ indicates that the assignment for the i-th object is not included in the counts. We treated the hyperparameters α, β^E and γ as constants in the model (we set α= 0.1, β^E= 0.000001, and γ=0.3).

We applied the dual route topic model to a small number of images from the LabelMe image set. We selected a set of 10 images to correspond with the 10 images used in the memory experiment. The images used in the simulation were selected based on having a relatively large number of annotations (30-60).

Up to this point, the model specifies a retrieval probability $p_i^{retrieve} = p(o_i \mid s)$ for each object i. Ideally, one would recall objects from this distribution strictly in order of decreasing probability. However, we assume that people cannot determine the strict order of probabilities. Therefore, we incorporate noise in the recall sampling process by letting the actual recall probability be based on a soft-max sampling process:

$$p_i^{recall} = \exp(\frac{1}{\tau} p_i^{retrieve}) \Big/ \sum\nolimits_j \exp(\frac{1}{\tau} p_j^{retrieve}) \qquad (8)$$

where τ is the parameter that controls the sampling noise. We set τ =0.008. In the experiment, participants were not allowed to repeat previous answers. To simulate this with the model, we sampled objects without replacement from the recall distribution.

To simulate the effect of study time we selected two subsets of the annotation word list for the images: a set of 80% of the annotations and a set of 20 % of the annotations. This corresponds to the idea that when studying an image for a restricted period of time not all the objects in the image are noticed. Subsets were created by drawing a random sample of objects

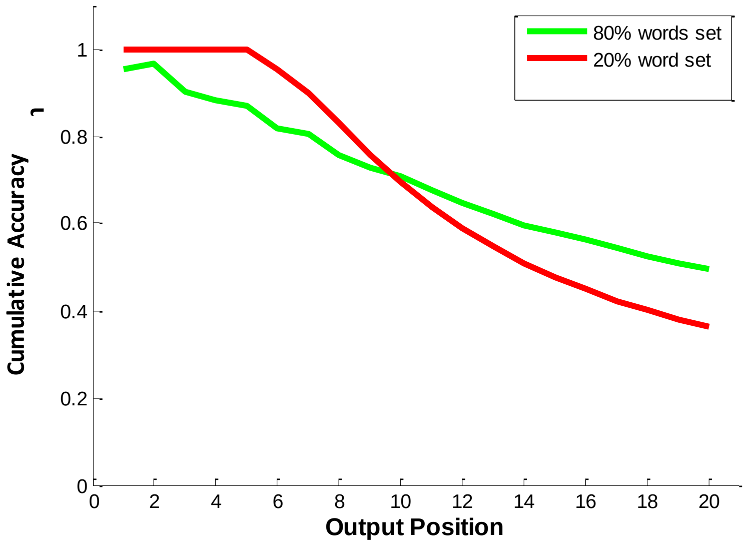


Figure 3.4: Model predictions: cumulative accuracy by output position when 80% and 20% of the objects have been perceptually encoded. The two conditions simulate the effect of long and short study times respectively.

from the full object set. Figure 3.4 shows the model predictions plotted in the same way as the results of our empirical study. The results show a qualitative fit to the experimental data. Objects from the smaller subset, corresponding to short study times, have initial higher accuracy, while a larger subset has initial lower accuracy followed by a cross-over. This models the somewhat counter-intuitive finding of our empirical study that the first few objects recalled for short study times are more likely to be correct than for longer study times. The model explains this finding because of different weightings of the two encoding routes, episodic and semantic. If a scene is studied for a longer period of time more objects are noticed and encoded, but it is more difficult to accurately store the longer list of object in memory because of the sparsity constraint in the episodic memory route. This leads to a greater number of objects encoded by the semantic route. While this route cannot fully reconstruct the objects present in the image, it is able to “guess” a larger number of objects, leading to relatively higher cumulative accuracy for later output positions. In contrast, seeing a scene for a shorter period of time, leads one to notice fewer objects but these objects can be encoded more effectively by the episodic route. However, the semantic context is not as well encoded in this case, leading to poorer performance for later output positions. Figure 5 show the probabilities of route assignments for three conditions: full set of objects, and the 80% and 20% subset conditions corresponding to long and short study times. Smaller word sets lead to greater episodic contributions, while larger word sets lead to almost equal contributions of episodic and semantic encoding routes.

The relative contribution of episodic and semantic information in recall can also account for other standard memory phenomena, such as the semantic isolation effect (von Restorff, 1933). An object is more likely to be recalled when it is part of a list where it violates the semantic context than when it is presented in a list where it is congruent with the semantic context. This finding can be explained by the dual route model because the route assignment to episodic and semantic memory routes is dependent on the context. Objects consistent with a scene (e.g., typical kitchen objects in a kitchen) can be explained by the semantic route, whereas an object that is not part of the semantic context of the scene (e.g., a ‘tree’ in a kitchen) can be explained by the episodic route assignment.

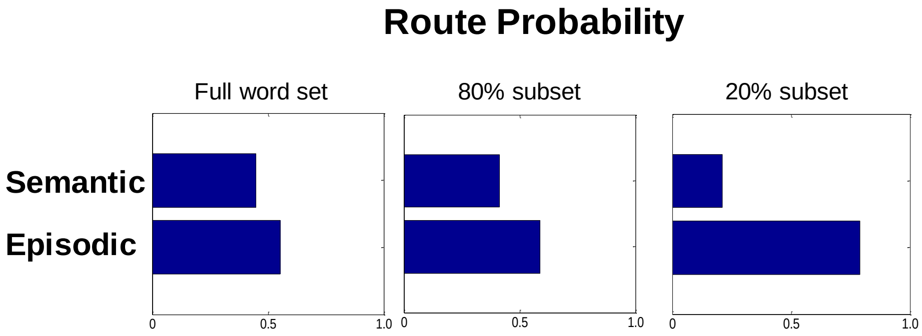


Figure 3.5: Model predictions for the full response set and for two sub sets of 80% and 20% of responses respectively.

To simulate the semantic isolation effect we created an artificial image where we manually determined the presence of objects. We selected an object that had an average recall probability within its semantic context – a tree in an urban scene (these are the same 10 scenes used in the previous two simulations). Figure 6 shows that 'a tree in an urban scene' is recalled with a slightly lower probability than the mean recall probability of all other objects in the scene. We then placed the 'tree' into a semantic context where it did not fit (e.g., a kitchen) by randomly removing an annotation in each of 10 kitchen scenes and replacing that annotation with 'tree'. The urban and kitchen scenes were equated for the number of annotations. We set α= 0.1, β^V= 0.01, and γ=0.3

Figure 6 shows the recall probability for the target object 'tree' and mean recall for all other objects on the list. Recall was higher for the target word than for the other objects. This is consistent with the finding for semantic isolation effects, as well as the idea that objects incongruent with the semantic context of a scene are recalled using episodic information.

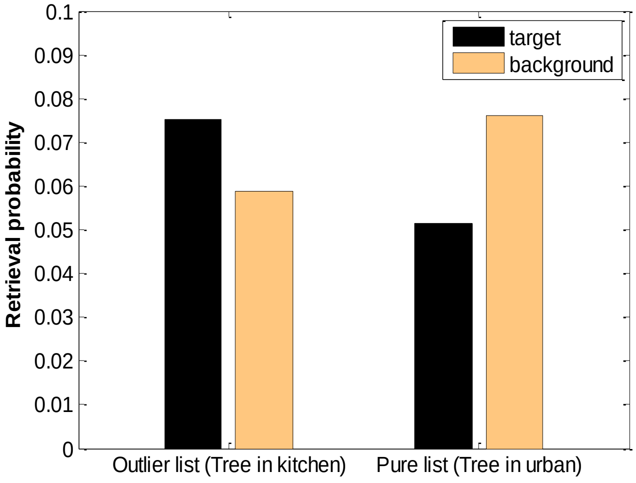


Figure 3.6: Model predictions for an object that is either incongruent (a tree in a kitchen) or congruent (a tree in an urban scene).

Conclusion

We have given an account of reconstructive memory, where reconstruction of objects in a scene is based on a mix of episodic memory traces and semantic context. Short study times lead to recall guided by episodic memory, whereas recall after longer study times is more influenced by semantic information. This counter-intuitive notion that longer study times lead to less reliance on episodic memory, is consistent with our empirical data showing that longer study times lead to an initially lower performance followed by a cross-over in accuracy. Given a dual route topic model account of reconstructive memory, where recall probability is given by the ability of an encoding route – episodic or semantic - to explain the occurrence of an object in a scene, this is to be expected. The model can also account for semantic isolation effects by favoring episodic encoding for objects that are not consistent with the semantic context of a scene.

Chapter 4

The Wisdom of Crowds with Informative Priors

Pernille Hemmer, Mark Steyvers & Brent Miller (2009). *Proceedings of the 32st Annual Conference of the Cognitive Science Society.*

Abstract

In some eyewitness situations, a group of individuals might have witnessed the same sequence of events. We consider the problem of aggregating eyewitness testimony, trying to reconstruct the true sequence of events as best as possible. We introduce a Bayesian model which incorporates individual differences in memory ability, as well as informative prior knowledge about event sequences, as measured in a separate experiment. We show how adding prior knowledge leads to improved model reconstructions, especially in small groups of error-prone individuals. This Bayesian aggregation model also leads to a "wisdom of crowds" effect, where the model's reconstruction is as good as some of the best individuals in the group.

Introduction

Studies of eyewitness testimony have shown that human memory can be incomplete and unreliable (e.g., Loftus, 1975). In real world situations, there might be multiple eyewitnesses, all of whom witnessed the same set of events. This raises the possibility of recovering the true account of events by analyzing the similarities in the recalled memories across individuals. Different individuals might also recall different aspects of the events, such that an aggregate narrative, based on the group's memory, would be closer to the true sequence of events than that of any one individual. An investigator might try to manually reconstruct the aggregate narrative, or witnesses might be allowed to discuss the events in order to develop the group narrative. Communication between witnesses however, has been shown to lead to much worse performance (Gagnon and Dixon, 2008), and humans have been shown to be inconsistent in assessing group information from multiple sources (Stasser & Titus, 1985). To avoid these problems, we propose a model of aggregation that can integrate the recalled memories from a number of independent individuals, while also taking in other important factors, such as individual differences and prior knowledge, into account.

Research on the "Wisdom of Crowds" (WoC) has shown that an aggregation of independent judgments often leads to a group estimate that is closer to the ground truth than that of most of the individuals (Surowiecki, 2004). These group estimates are often simply found by taking the mean, median, or mode of responses (Galton, 1907; Surowiecki, 2004). Much of the previous literature on aggregation of judgments has focused on tasks where individuals estimate numerical quantities and probabilities (Budescu, Yu, 2007; Hogarth, 1978; Wallsten, Budescu, Erev, & Diederich, 1997). It is, however, often that case that eyewitness have to retrieve information more complex than single numerical estimates.

The WoC effect can also be demonstrated with more complex problem sets. For example, the WoC effect has been demonstrated with solutions to problem-solving situations such as finding minimum spanning trees for a set of nodes (Yi, Steyvers, Lee & Dry, in press). Steyvers, Lee, Miller, and Hemmer (2009) showed that order information from semantic memory can also be combined across individuals to give high accuracy in reconstructing the true order of items along some physical or temporal dimension; when individuals recalled the order of US presidents, or the order of rivers according to length, many of the individual orderings were error-prone, but the aggregate orderings were more accurate, on average. In Steyvers et al. (2009), a number of aggregation models for order information were tested. It was found that using Bayesian models that incorporated psychologically plausible representations, cognitive processes and individual differences outperformed basic heuristic aggregation approaches, such as taking the mode.

When errors across individuals are uncorrelated (as they tend to be when individuals independently give their judgments) the errors will cancel out in the aggregate. Therefore, one expects the best results in WoC experiments with a large number of individuals. In eyewitness situations however, there is rarely a "crowd" available to witness the same set of events. In these cases, we have to rely on a small number of individuals (in many cases, just one) and significant errors might not cancel. Therefore, it might not be sufficient to just analyze the commonalities across the witness reports. We propose that it is better to combine the witness reports along with prior knowledge about the particular event sequence. Combining prior knowledge with noisy information has been shown in other domains to improve the recovered estimate (Hemmer & Steyvers, 2008; Konkle & Oliva, 2007; Kan, Alexander, Verfaelle, 2009).

We focus in this research on the problem of reconstructing event sequences. The goal is to reconstruct the true ordering of a set of events by aggregating the recalled orderings from a small number of individuals, all of whom witnessed the same event sequence. The novelty of the current approach is that we incorporate informative prior knowledge in an aggregation model for order information in order to improve the aggregate estimate. This is especially helpful when aggregating across a small number of error-prone individuals.

We present our results as follows. We first report on behavioral experiments wherein we tested people's ability to reconstruct, from episodic memory, the order of stereotyped events (e.g., getting up in the morning), or random events (e.g., clay animation without a clear story line). We also report on experiments where we measured prior knowledge for the same set of events. We then describe a Bayesian approach that aggregates the orderings across individuals while taking prior knowledge into account.

Empirical study on serial recall

Much research on serial recall has been done on random word and letter sequences that do not have any obvious organization. In such experiments, individuals are shown a sequence of words or letters, and the task is to recall the original temporal order as best as possible during a later test. Typical errors in the recalled orderings are transposition errors where the orderings are locally perturbed (Estes, 1997; Nairne, 1992) -- two events nearby in time tend be reconstructed as occurring nearby but the amount of perturbation noise depends on many factors such as time elapsed between study and test, stimulus characteristics and individual differences. Similar patterns have been observed in more naturalistic experiments, such as naming the day of the week an event occurred (Huttenlocher, Hedges, & Prohaska, 1990), as well as for autobiographical memory, such as ordering the events of September 11th (Altmann, 2003). With more naturalistic event sequences, prior knowledge about the event sequences can influence episodic memory. People have clear expectations for routine activities and are sensitive to the ordering of actions within an activity (Bower, Black & Turner, 1979).

We conducted a series of behavioral experiments using two types of event sequences. We used a number of *stereotyped* event sequences, such as getting up in the morning, or jumping on a bus, for which people have clearly defined expectations, and a number of *random* event sequence, such as clay animation sequences or Japanese pizza commercials, for which the temporal organization might be less structured. To assess the prior knowledge people have about these types of events, we first conducted a prior knowledge study where we asked participants to order the events in the most natural order possible without actually showing them the original, true event sequence. This allows us to estimate a model for the prior probability of each sequence.

In a separate experiment, we assessed serial recall for each of event sequences. It should be noted that our definition of serial recall differs from the standard use of the term in that our task only involves ordering the events, not recalling the items to be ordered, as in a standard serial recall task. In our task, we first showed a video of the original event sequence which was followed by a serial recall test in which individuals ordered image stills from the video as best as possible according to the original temporal sequence in which the events appeared. No communication between individuals was allowed in any of our tasks, and therefore the data consists of independent recollections from individuals.

prior

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4	3	13	1	9	5	3	12	11	10	10	2	5	1	1	10
A	B	A	A	A	B	B	A	B	B	A	A	D	B	A	B
B	C	C	B	C	D	A	D	A	A	C	B	B	A	B	A
E	A	E	C	D	A	D	G	C	E	E	D	A	C	D	E
D	D	F	D	E	E	C	C	E	C	D	C	C	D	C	F
C	E	G	E	G	C	E	E	D	G	G	E	E	E	E	D
G	G	H	G	F	G	G	F	J	F	F	G	G	F	F	G
F	F	I	F	H	F	F	H	I	H	H	F	F	G	G	H
H	H	D	H	I	H	H	I	H	J	I	H	H	H	H	I
I	I	J	I	J	I	I	J	F	I	J	I	I	I	I	J
J	J	B	J	B	J	J	B	G	D	B	J	J	J	J	C

serial recall

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
2	1	2	4	2	5	1	1	1	1	2	1	6	1	1	0	1	2	2	1	2	1	1	1	2	0	1	1
B	A	A	A	B	B	A	A	A	A	A	A	A	A	A	A	A	B	B	A	A	A	B	A	B	A	A	A
A	B	B	B	A	A	B	B	B	B	B	B	G	B	B	B	B	A	A	B	B	B	A	C	A	B	B	B
C	C	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	B	C	C	C	C
D	D	D	C	D	D	D	D	D	D	D	D	B	D	D	D	D	E	D	D	D	D	D	D	D	D	E	D
E	E	G	D	E	H	E	E	F	F	F	E	D	F	E	E	E	D	E	E	F	E	E	E	E	E	D	E
G	G	E	E	G	F	G	G	E	E	G	G	E	E	G	F	G	F	G	G	G	G	F	F	G	F	F	G
F	F	F	F	F	E	F	F	G	G	E	F	F	G	F	G	F	G	F	F	E	F	G	G	F	G	G	F
H	H	H	H	H	G	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I
J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J	J

Figure 4.1. The sequence A-J shows the 10 images from the 'bus' video sequence in the correct temporal order. The two tables show the participant orderings in the prior knowledge and serial recall experiment. The first row is the participant id. The second row is the Kendall's tau distance between the true ordering and the recalled order for that participant.

Methods

Participants were undergraduate students at the University of California, Irvine. There were 16 participants in the prior knowledge experiment and 28 participants in the serial recall experiment.

Materials

We sampled 6 videos from YouTube.com. Three videos depicted stereotyped events sequences (getting up in the morning, a wedding, getting on the school bus). Three videos depicted more random event sequences (a Japanese yogurt commercial, a Japanese pizza commercial, and a clay animation sequence). For each of the 6 videos 10 still images of individual scenes were drawn. See Figure 4.1 for an example.

Prior Knowledge Experiment

Participants were shown 10 image stills from a given event sequence (e.g., Wedding) and asked to order the 10 images based on their prior expectation of how the event in the slides might unfold. Importantly, in this experiment, participants were never shown the original video sequence from which the image stills were drawn. They responded using an interactive interface in which the images were randomly ordered on the screen and the instruction was to order the images in any way to make the sequence as natural as possible.

Serial Recall Experiment

Participants first viewed the original video sequence. Participants were then presented with the same interface as in the prior knowledge experiment. They were shown 10 image stills that they had to order in the original temporal order. For both the prior knowledge and memory experiment, the initial ordering of the 10 image stills, as well as the order of the 6 video sequences, was randomized across participants.

Results and Discussion

To evaluate the performance of participants, we measured the distance between the reconstructed and the correct ordering. A commonly used distance metric for orderings is Kendall's τ (Marden, 1995). This distance metric is the minimum number of adjacent pairwise swaps necessary to resolve any disagreements between the two orderings being compared. Values of τ range from $0 \leq \tau \leq (N-1)/2$, where N is the number of items in the order: $N=10$ for all of our event sequences. In our experiment, a $\tau = 0$ indicates that the participant responded with the exact correct ordering. A $\tau = 1$ indicates that one adjacent pair of items was swapped. When participants are using a random guessing strategy, their expected mean expected distance is $\tau = (N-1)/4 = 22.5$.

Figure 4.1 shows the raw data collected for the "bus" video sequence – a stereotyped event sequence. In the prior knowledge experiment, participants produced orderings that were much better than chance, suggesting that a priori, it is possible to guess the true ordering of events in these types of event sequences. In the memory experiment, 2 participants produced the correct ordering, and 15 more were within one swap of the true order. Note that very few identical orderings are produced between participants. We found that for all 3 random events, in both the prior knowledge experiment and the memory experiment, each participant produced a unique ordering. For the 3 stereotyped event sequences however, only one sequence led to unique orderings across all participants.

Figure 4.2 shows the distributions of the Kendall τ distances for the serial recall and prior knowledge experiment. The top panel shows the distances for stereotyped event sequences and the bottom panel shows the distances for random event sequences. The dashed line shows the distribution of distances that can be expected from a random guessing strategy (this distribution can be calculated exactly, see Marden, 1995). For both the stereotyped and random event sequences, the distances are lower for the memory task than for the prior knowledge task. The distances are also lower for the stereotyped event sequences than for the random event sequences. Even when participants did not study the videos (the prior knowledge condition), they performed better than chance in the stereotyped condition, as compared to the random condition where prior knowledge performance led to a distribution of distances very similar to distances expected from chance performance. These results demonstrate that general knowledge about events can greatly contribute to the accuracy of recalling these events.

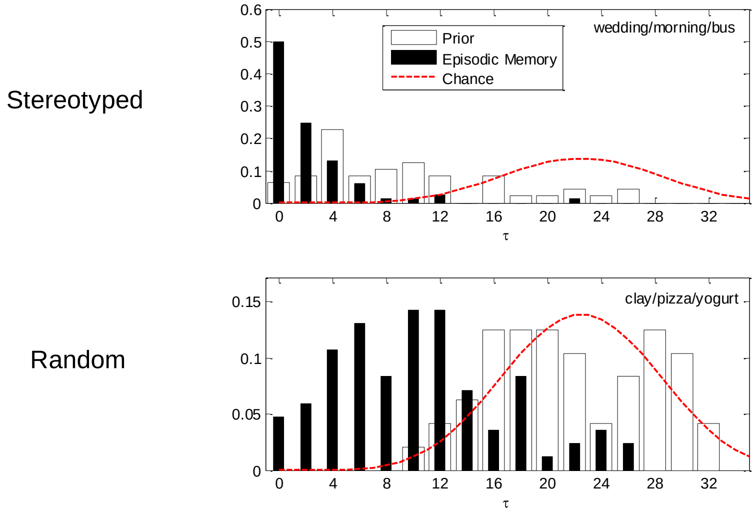


Figure 4.2. Distributions of Kendall τ distances.

Modeling

We can conclude from our empirical study that prior knowledge can lead to improved average performance in recall. When ordering scenes from an event with strong prior expectations, the resulting orderings are relatively close to the true ordering. Of course, performance improves on average after observing the true event sequence and later recalling the sequence from memory. This raises the question of how one might incorporate an informative prior in a model for aggregating rank-ordered recall. Such priors might guard against errors from a small number of poorly performing individuals. In this paper, we explore very simple models to aggregate the orderings of individuals. The goal of the modeling is not to build a comprehensive model of recall that specifies all the representations and processes involved in storing and retrieving information from memory. Instead, we will focus on simple probabilistic models such as a Mallows model (e.g. Steyvers et al., 2009) that allow us to aggregate the retrieved orderings from a number of individuals using Bayesian inference. The current model incorporates two important differences to the previous work by Steyvers et al. (2009). First, we generalize the model to allow for individual differences in memory performance. These individual differences are estimated by the model in a purely unsupervised fashion and do not require knowledge of past performance in other tasks or access to a known ground truth. With the individual differences, the model finds aggregates that are weighted towards solutions provided by the individuals that are estimated to have good memory performance.

Second, we develop a simple extension of Mallows models that allows for informative priors. This prior is estimated from the orderings produced in the prior knowledge experiment.

Mallows Model with an Uninformative Prior

In a basic Mallows model (Marden, 1995), all individuals are assumed to derive their orderings from a single underlying ordering, that we will refer to as the *group knowledge*. The group knowledge is a latent variable in the model that can be estimated from the data. Importantly, Mallows model assumes that each individual produces orderings centered on the group ordering with distant orderings less likely than orderings close to the group ordering. Although Mallows-type models have often been used to analyze preference rankings (Marden, 1995), they have not been applied, as far as we are aware, to ordering data from serial recall experiments. In our first extension of the standard model we allow for individual differences in memory performance. We evaluated this aggregation model by comparing the estimated group ordering to the ground truth. If the model is able to tap into the collective wisdom of a group of individuals, the estimated group ordering should be close to the true ordering.

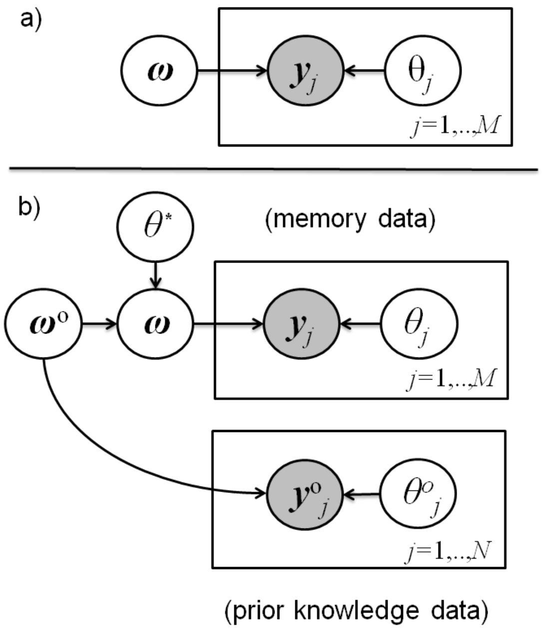


Figure 4.3. The graphical model representations for the Mallows model with an uninformative prior (a) and an informative prior about the group knowledge (b).

Specifically, let $\boldsymbol{y}_j$ represent the ordering from individual j, and $\boldsymbol{\omega}$ the latent group ordering. In a Mallows model, the probability of each individual ordering given the group ordering is given by

$$p(\boldsymbol{y}_j|\boldsymbol{\omega},\theta_j) \propto e^{-\mathrm{d}(\boldsymbol{y}_j,\boldsymbol{\omega})\theta_j} \tag{1}$$

where for simplicity we have omitted the normalization constant. The function d returns the Kendall τ distance between two orderings. The scaling parameter θ_j determines how close the observed order for individual j is to the group ordering. It can be interpreted as an individual (inverse) noise parameter -- good individuals tend to closer to the group consensus (high θ) whereas poor performing individuals return more idiosyncratic orderings further away from the group knowledge (low θ). We will assume a Gamma prior on the individual noise levels: $\theta_j \sim \mathrm{Gamma}(\theta_0\lambda,\ 1/\lambda)$, where λ is a hyperparameter that sets the overall level of cohesion expected from the group. Notably, in this first model, we have assumed a uniform prior over group orderings, $\boldsymbol{\omega}$~Uniform(Ω), where Ω is the set of all orderings. Therefore, a priori, the model assumes no preference for a particular group ordering.

Figure 4.3, panel a, shows a graphical representation of the model. Shaded nodes represent observed variables while nodes without shading represent latent variables. The arrows indicate the conditional dependencies between the variables and the plate represents the repeated sampling steps across *M* subjects in the memory experiment.

Mallows Model with an Informative Prior

We now introduce a simple variant of this model that allows for an informative prior. The idea is that the group knowledge is itself sampled from a Mallows model:

$$p\left(\boldsymbol{\omega} \middle| \boldsymbol{\omega}^{\mathbf{0}}, \theta^{*}\right) \propto e^{-\mathrm{d}\left(\boldsymbol{\omega}, \boldsymbol{\omega}^{\mathbf{0}}\right)\theta^{*}} \qquad (2)$$

where $\boldsymbol{\omega}^{\mathbf{0}}$ is the prior ordering from which the group ordering is derived, and θ^{*} is a scaling parameter. This prior stage in Mallows model at first might not seem to gain any additional information because it is not clear how the prior ordering can be constrained. However, we have data in the prior knowledge experiment in which *N* participants tell us what orderings they expect from certain scenes. Let $\boldsymbol{y}_j{}^0$ represent the prior ordering given by individual *j* in the prior knowledge experiment. We assume that these are produced by a Mallows model with $\boldsymbol{\omega}^{\mathbf{0}}$ as the "center":

$$p\left(\boldsymbol{y}_j^0 \middle| \boldsymbol{\omega}^{\mathbf{0}}, \theta_j^0\right) \propto e^{-\mathrm{d}\left(y_j^0, \boldsymbol{\omega}^{\mathbf{0}}\right)\theta_j^0} \qquad (3)$$

Figure 4.3, panel b, shows the corresponding graphical model. With this model, we are setting a prior on the group ordering -- when there is only data available from a few individual in the memory experiment, the group ordering will be influenced by the data from the prior knowledge experiment leading to group orderings that are a priori deemed likely. When data from more individual becomes available in the memory experiment, the prior knowledge data will have a diminishing influence on the group ordering which will be mostly determined by the memory data.

Modeling Results

All latent variables in the model were estimated using a MCMC procedure, separately for each event sequence. The result of the inference procedure is a probability distribution over group orderings, of which we take the mode as the single answer for a particular problem. Note that the inferred group ordering does not have to correspond to an ordering of any particular individual. The model just finds the ordering that is close to all of the observed memory orderings.

Figure 4.4 shows the calibration for the two models on a single event sequence (the clay animation video). Each panel shows the relationship between the inferred θ (related to the distance of each individual to the group ordering) and the Kendall's τ distance of the individual's answer to the ground truth. The plots show that individuals who are close to the group ordering tend to be closer to the ground truth. This means that the models can calibrate the performance levels of individuals, even in the absence of any explicit feedback or access to the ground truth.

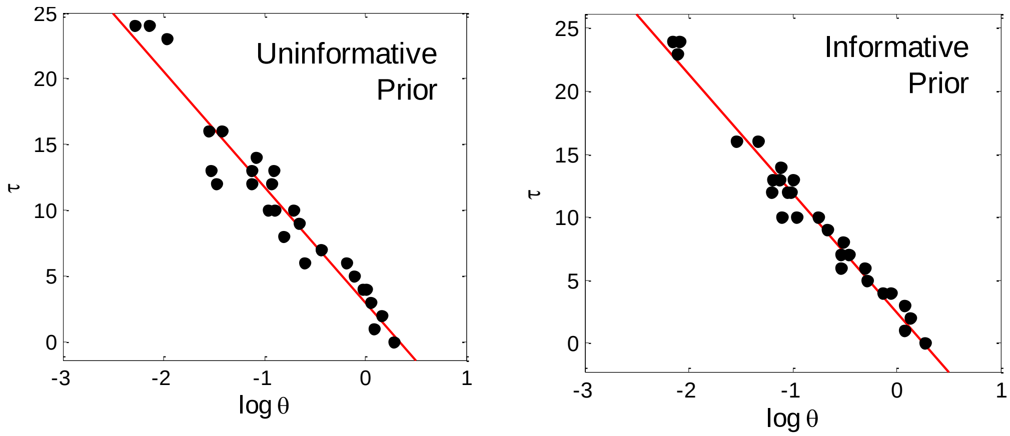


Figure 4.4. Calibration results for the two models for one event sequence.

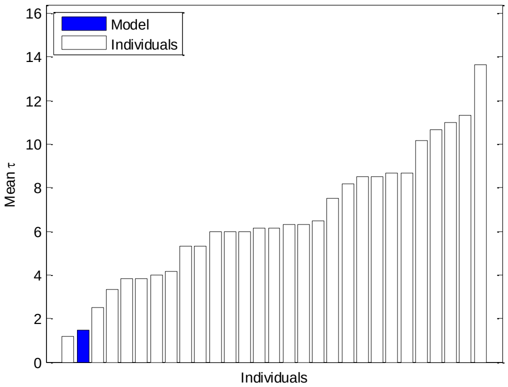


Figure 4.5. Performance of individuals and model (with informative prior) averaged over six event sequences.

Figure 4.5 shows the Kendall's τ distance for each individual in the memory experiment averaged over the six event sequences. Note that there are substantial individual differences with some individuals coming relatively close to the ground truth. The figure also shows the average model performance. Comparison between individual and model performance reveals a WoC effect: The model performs as well as some of the best individuals, with only one individual outperforming the model. Therefore, we can conclude there is a weak WoC effect (a strong WoC effect would correspond to a situation where the model outperforms all individuals in the group). We now focus on applying the model to subsets of participants to mimic eyewitness situations that typically involve only small number of individuals. In the first analysis, we select a random set of *K* individuals from the original set of 28 individuals. We then apply the two models to the subset of individuals. Figure 4.6 shows model results for the model with the informative and uninformative prior separated for stereotyped and random event sequences. For random event sequences, where the prior is weak, there is no improvement in the aggregation between the two models (if anything, there is a small performance decrement for the model with the informative prior). For stereotyped event sequences however, people have strong prior expectations about the true ordering of events and there is a marked improvement in the aggregate response in the

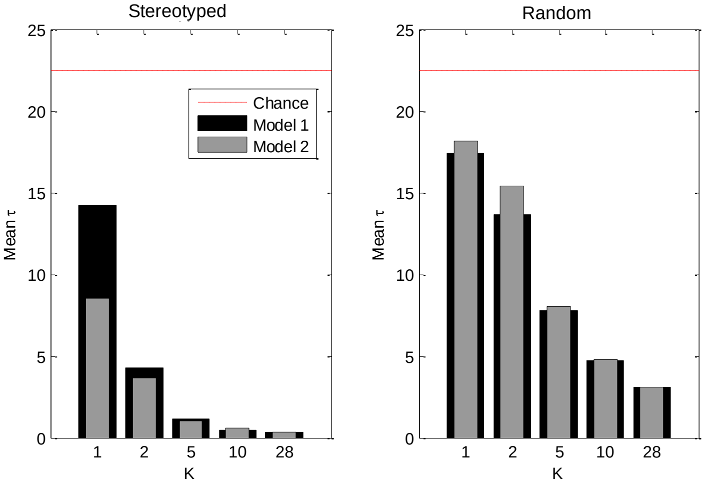


Figure 4.6. Results from the models with an uninformative prior (model 1) and informative prior (model 2) for *random* subsets of *K* individuals from the memory task.

model with the informative prior. This improvement is most pronounced with low sample sizes (*K*=1 and *K*=2) when the prior can still exert an influence on the inferred group orderings. Note that when *K*=1, the model with the uninformative prior has no information other than the ordering given by a single individual – therefore, the aggregate solution given by the model is equivalent to the ordering provided by the individual. This results in an average tau of around 15. However, performance for the model with the informative prior is much better resulting in a tau of around 8, because the aggregate solution combines the single remembered ordering with the a priori likely orderings.

To better highlight the benefit of the prior information, we also conducted a model analysis where we selected the *worst* performing individuals in the sample. In this sampling procedure, we sample the *K* worst individuals where we vary *K* from 1 (the single worst performing individual) to 28 (all individuals combined). Figure 4.7 shows model results for both models separated for stereotyped and random event sequences. The relative performance benefits

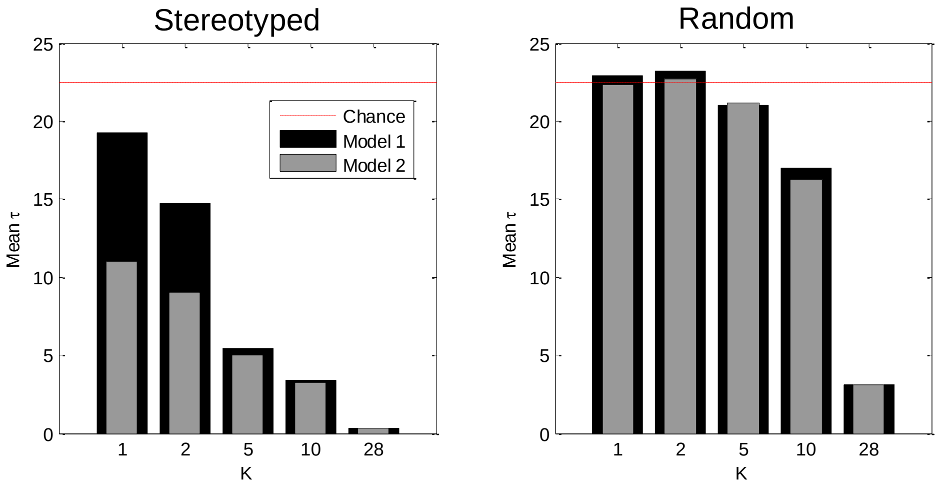


Figure 4.7. Results from the models with an uninformative prior (model 1) and informative prior (model 2) for subsets of the *worst K* individuals from the memory task.

can be seen most clearly for the stereotyped event sequences for low sample sizes (*K*=1 and *K*=2). In these cases, the worst individuals recall event sequences that are a priori unlikely and the prior "corrects for" the noise in the available data.

Therefore, these analyses suggest that an aggregation model with informative priors can be used to guard against the most egregious errors committed by the worst individuals in the memory task.

Conclusion

We have presented two approaches for aggregating recalled sequences of events in order to reconstruct the true event sequence as best as possible. Individuals are likely to differ in their ability to recall event sequences and pay attention to different parts on an event sequences. Therefore, by analyzing the consistencies in orderings across individuals, we can extract the collective wisdom in the group. We presented two aggregation approaches based on Mallows model that allow for individual differences. The models combine information at the group level with information at the individual level to explain orderings given by an individual. In the first approach, the model uses only the data from the individuals who all witnessed an event sequence. In the second approach, the model uses an additional source of data based on the prior knowledge about the events extracted from another group of individuals.

We demonstrated a weak WoC effect, where the average performance of the model was better than every individual, save one. We have also shown that a Mallows model with informative priors has a markedly improved ability to reconstruct the ground truth in cases where the event sequences are highly stereotyped and a small sample of poorly performing individuals is used for aggregation. This is particularly important in eyewitness situations where we typically have only a small number of individuals available.

Chapter 5

The Influence of Real World Prior Knowledge on Episodic Memory

Pernille Hemmer, Jenny Shi & Mark Steyvers (Submitted)

Abstract

Many aspects of our experiences do not have to be explicitly remembered, but can be inferred from knowledge of regularities in our environment. Such knowledge can operate at multiple levels of abstractions; recall for the height of a person can be influenced by general knowledge about heights of people, or by specific knowledge about the heights of men and women. We show that the a priori expectations people have of the heights of men and women are in close correspondence with the distribution of heights in the population. We also assess the contribution of prior knowledge on reconstructive memory. Memory performance is tested in a recall task where subjects reconstruct from memory the height of people shown earlier in a sequence. Our results suggest that prior knowledge can improve average recall, and that knowledge can come from multiple levels of abstraction such as gender and the overall heights of people.

Introduction

Imagine that you witnessed a crime committed by two people – a man and a woman. Later, you are asked by the police to recall certain physical features of the suspects, such as their height. To reconstruct the event that you witnessed, you may use not only episodic information related to the specific incident, but also general knowledge that you have about peoples' physical characteristics. For instance, you might infer that the woman involved was about 5'4" (163cm) – not because you remember her height exactly, but because you know that this is the average height of women in general. When asked about the male, you might infer that he was about 5'8" (173cm) based on what you know about the average height of men. Using knowledge of the regularities of our environment, many aspects of our experiences do not have to be explicitly remembered, but can be inferred based on this knowledge. Regardless of whether such supposition is deliberate or unconscious, prior knowledge can have a significant influence on recall.

Drawn from natural observations and experiences, people have a general expectation that discernable differences exist between males and females. These discrepancies may be physiological, such as height (e.g., Biernat, 1993), or behavioral, including cognitive patterns and social roles (e.g., Eagly & Steffen, 1984). Our perceptions and memories are influenced by our awareness of the true distributions we observe in nature (Konkle & Oliva, 2007). Some aspects of natural observations might be so salient and ubiquitous that people develop deep-seated beliefs that cannot be ignored; e.g., that gender is predictive of a person's height. As evidence of this, such biases can affect decisions even when people are informed to disregard their prior expectations. For example, when asked to judge height without regard to gender, people could not omit their prior knowledge of gender differences as it relates to height (Nelson, Biernat & Manis, 1990). People are also quite accurate when asked to make perceptual judgments based on knowledge of natural distributions, e.g., estimating height based on accessible gender information (Kato & Higashiyama, 1998). Applying this to the eyewitness example, it is expected that knowing the gender of the crime suspect has an impact on estimating their height. Yet, the question is: Does the use of prior knowledge lead to a more accurate estimate because additional information is taken into account, or does it hinder performance because stereotyped information can lead to systematic biases?

Prior knowledge about events is often portrayed as the cause of errors in recall. A recurring feature in the false memory literature is that events that are more consistent with prior expectations result in greater intrusions in recall. When presented with stimuli that have strong associations, which are deliberately withheld by the experimenters, participants are more prone to falsely recall these targets (e.g., Roediger and McDermott, 1995, Brewer and Treyens, 1981). Common scenarios have also been shown to result in greater levels of intrusions than uncommon scenarios, e.g., children are more likely to falsely recall having their finger caught in a mousetrap compared to having an enema– presumably because the first is more familiar (Otgaar, Candel, Scoboria and Merkelbach, 2010).

On the other hand, prior knowledge can also be beneficial to memory. The availability of meaningful information can improve recall for abstract concepts (Bartlett, 1932). Having knowledge about the prior distribution of the stimuli can also lead to an average improvement in recall (Huttenlocher et al., 1991). Memory for object size is naturally biased towards the normative size of objects consistent with their real-world size (Konkle and Oliva, 2007). Using prior knowledge might be a prudent strategy, because it provides a natural bound on errors. 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. This is true even when participants do not remember studying the objects (Hemmer and Steyvers, 2009a; 2009b). The key of these experiments is that the stimuli are consistent with the environment from which they are drawn and representative of the event to be remembered.

In the current study, we propose to evaluate the influence of prior knowledge on recall as it pertains to height and gender. Our use of gender and height distributions as prior knowledge is innovative in its approach. We use a novel method to evaluate reconstructive memory based on stimulus distributions that are representative of environmental distributions. Previous studies that investigated similar effects using gender and height employed different strategies and intended to assess social biases through judgment tasks rather than memory tasks. For example, judgments of human height allowed the measurement of base rate beliefs and the assessment of stereotype resilience across developmental age groups (Biernat, 1993). Judgment tasks of height and gender have also been used to examine the generality of belief-biased reasoning by assessing the relationship between cognitive ability and flexibility in prior belief (Sa et al., 1999). The commonality of these studies is that they use images of males and females and height estimates were made as either paired comparison judgments or numerical expressions. To the best of our knowledge, there have been no previous memory studies using naturalistic images of females and males as stimuli. One memory study assessed the veracity of eyewitness recall for height by examining a database of witness statements of real crimes (Fahsing, Ask and Granhag, 2004). While this study involved naturalistic stimuli, it was not a controlled memory experiment. Another study evaluated recall for group characteristics (such as the proportion of individuals above a certain height) as a function of individual height traits (Rothbart, Fulero, Jensen, Howard and Birrell, 1977). Here the stimuli were presented in the format of sentence statements. Although this study was a controlled laboratory experiment, it did not use naturalistic stimuli.

Using height and gender as stimuli has several advantages. First, there is a prominent difference in the mean height between genders, which allows recall to be evaluated hierarchically – height can be relative to the overall height in the population or relative to the specific height of men and women. The influence of prior knowledge has been shown to operate on recall at various categorical levels. For example, Hemmer and Steyvers (2009a; 2009b) found that prior knowledge had different effects at multiple levels of abstraction, and proposed that the influences are hierarchically structured. For instance, objects with limited categorical information (artificial shapes) were biased towards the mean of the overall distribution of artificial shapes, whereas objects with clear categorical information (fruits and vegetables) were 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). A hierarchical influence of prior knowledge makes two clear predictions about the effect of prior knowledge on episodic memory. First, that there will be an overall regression to the mean, where recalled height for people at heights below the mean population height will be overestimated while recalled height for people at heights above the mean population height will be underestimated. (See Figure 5.1, Panel A). The second and critical prediction is that when two people (a short male and a tall female) are studied at the exact same height, recall will be differentially biased towards the height distributions that are gender specific. The tall female will be underestimated towards the mean of female height and the short male (objectively the same height as the female) will be overestimated towards the mean of male height. In other words, prior knowledge will differentially affect the memory of two people

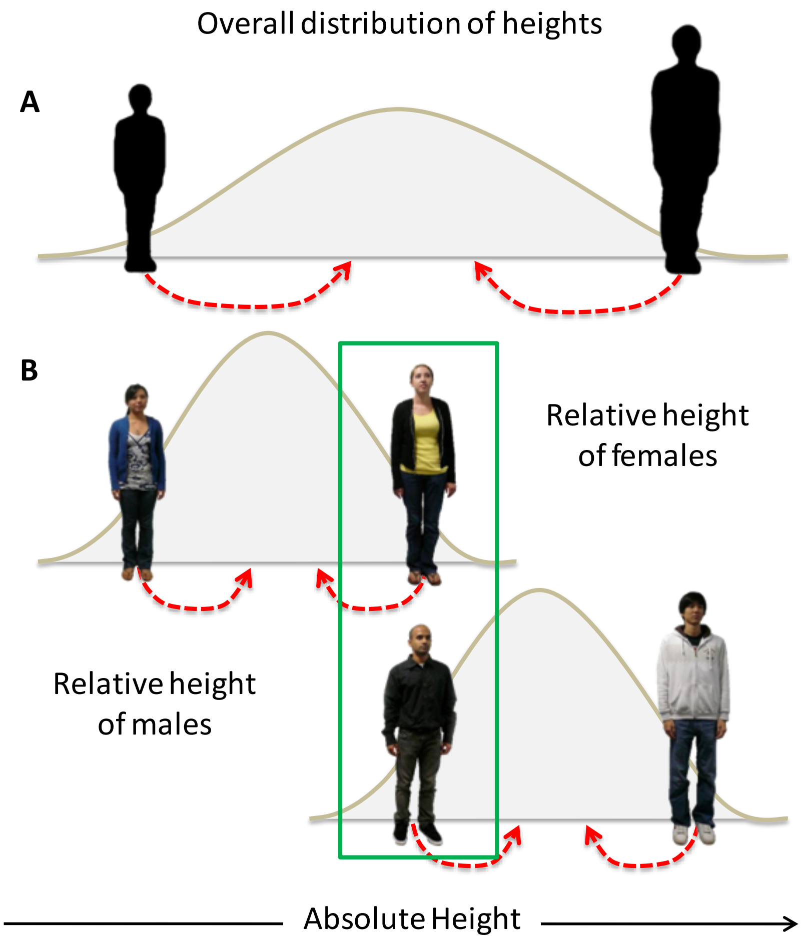


Figure 5.1. Predicted biases in recall as a function of A) limited gender information, and B) clear gender information for females and males.

originally presented at the same height. Figure 5.1, Panel B illustrates this effect. Thus, prior knowledge at a more fine-grained level might contribute to further improvements in average recall over general level knowledge.

A further advantage of using height and gender is that there is an explicit ground truth that can be measured directly. The critical difference between the current approach and that of Hemmer and Steyvers (2009a; 2009b) is the way in which the "true" size of the stimulus is obtained. Hemmer and Steyvers did not directly measure the size of fruits and vegetables. Rather, they used a norming study to ascertain the size ranges of their stimuli by asking people to make judgments about the possible sizes of the objects (smallest, largest and average size). While they maintained a focus on more naturalistic stimuli, the sizes of fruits and vegetables can vary quite broadly across cultures and time. For instance, where watermelons may only be experienced as round in the United States, they are available in square and even heart-shapes in other countries. Although there was a large degree of agreement between participants in the normed size judgments, it is possible that semantic inconsistencies in the stimuli affected the results, thereby diminishing the efficiency and salience of prior knowledge. In the current study we seek to test the influence of prior knowledge for the height of people. To obtain the ground truth and the size ranges used in the memory experiment, we will measure the height of people directly and not rely on the psychological judgments of participants.

In two separate experiments, we investigated the interaction between prior knowledge and reconstructive memory. In Experiment 1, we measured performance in a reconstructive memory task where participants study images of the heights of females, males and gender ambiguous silhouettes. In Experiment 2, we assessed people's prior knowledge for the same stimuli.

Experiment 1

Participants

Twenty-two undergraduate students at the University of California, Irvine participated in the experiment. The participants were not involved in the stimulus development phase and were compensated with course credit.

Stimuli

We developed naturalistic stimuli in the form of photographs of real people. We photographed 212 randomly selected male (68) and female (144) students at the University of

California, Irvine. They participated in exchange for course credit. All images were taken against a white wall and next to a blue door. The subjects were required to stand up straight, have their hands to the side, maintain a neutral expression and look at a designated fixation point on the opposite wall. Women were required to have their hair up. All participants were instructed to stand in a fixed position relative to the camera. The distance from the camera to the participants' heels was exactly 230 cm. The camera was maintained at 120 cm of elevation from the floor using a tripod. Each individual was measured and their height was recorded in whole inches. See Figure 5.2 for sample images.

To ensure that our sample of individuals was representative of the general population, we compared our sample distribution to height statistics obtained from the Center for Disease Control (McDowell et al. 2008). Our sample ranged in height from 147.3–182.9 cm for females and from 162.6 – 193 cm for males. The sample was normally distributed around a mean of 162.9 cm for females and 175 cm for males. This is comparable to the range and distribution in the US population over age 20 with a mean of 162.1 cm for females and 176.3 cm for males. Figure 5.3 shows the cumulative density function for the stimuli and the CDC data.

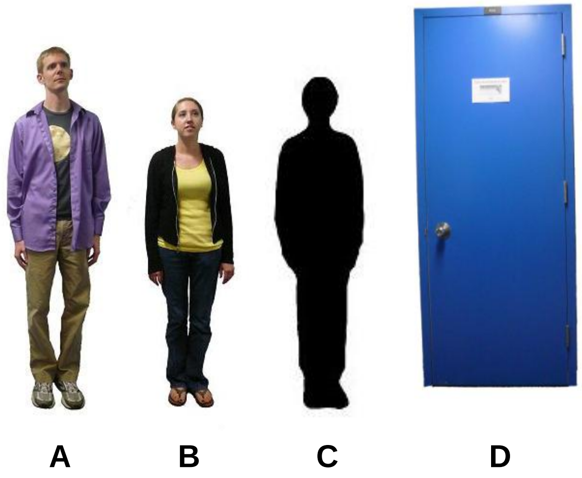


Figure 5.2. Stimulus examples from the prior knowledge and memory experiment. A) an unmasked male, B) an unmasked female, C) an ambiguous silhouette, and D) the reference image of a door.

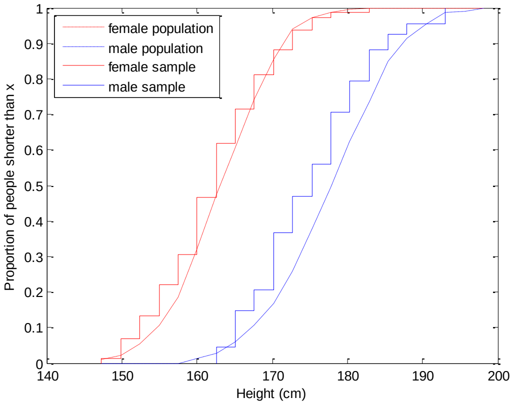


Figure 5.3. Cumulative density functions for the population and stimuli by gender. Red indicates female, blue indicates male. Solid lines represent stimuli; dashed lines represent the true population (obtained from the CDC).

From the 200 photographs we selected 48 images, 24 female and 24 male, to be used in the two experiments as our ‘unmasked’ images. Importantly, the selected images for each gender were distributed with the same range and frequency as the heights in the general population based on the CDC data. The purpose of the unmasked stimuli was that it retained all aspects of the figure, allowing prior knowledge of gender to be accessed (For example stimuli see Figure 5.2, panel A and B).

We also created a category of ‘masked’ images to be used as a comparison in the memory study. To produce the masked stimuli, we filled the interiors of the 48 unmasked stimuli with a black color. The masked stimuli retained only the height and human figure information, which strongly limits the accessibility of gender information (See Figure 5.2 panel C for a sample of the masked stimuli). To ensure that the stimuli were adequately ambiguous and that the physical characteristics were not predictive of height judgments, we asked ten individual to guess the gender of the stimuli. Based on these evaluations we selected 24 images, 12 of each gender, that were the most ambiguous and most representative of the true height distribution to use in our memory experiment. All unmasked stimuli measured 456 x 1229 pixels. All masked stimuli measured 285 x 768 pixels.

Procedure

To assess how prior knowledge of gender influences recall for height, we implemented a continuous recall paradigm where the study and test phases are randomly intertwined. Seventy-two images were blocked by two categories (unmasked and masked). Block 1 consisted of 48 unmasked stimuli (24 male and 24 female) and Block 2 consisted of 24 masked stimuli. An image of a door was displayed as a comparison image on the left side of the screen (See Figure 5.2, D). Prior to the experiment, the participants were shown the actual door that was used as a reference in the task. Each study image was presented for 2 seconds on the right side of the screen and was presented at the true height relative to the door. Upon presentation of test trials, the height of the stimulus was initialized at a random size. At test, participants were asked the following question, "What was the height of this person, compared to the door on the left, when you saw them at study? If you are not sure, make a best guess." To reconstruct the studied size, the participants used the computer mouse to move a slider on the right edge of the screen. Once they had scaled the figure to the size they recalled from study, they clicked on a button labeled "OK" and proceeded to the next trial until the experiment ended. Trials were randomized across and within blocks.

Results

To investigate the effect of prior knowledge on recall for the height of males, females and ambiguous silhouettes, we measured recall error as the difference between the recalled size and the studied size. All measurements will be given in centimeter (cm) units.

Figure 5.4 shows the recall errors as a function of absolute study size for each category – female, male and masked stimuli. The lines in the figure have two important aspects: slope and intercept. We first evaluated the bias toward the mean height. A negative slope indicates a bias to the center of the range of heights, such that short people are overestimated and tall people are underestimated.

The observed pattern of correction toward the category center as indicated by negative slopes for all categories supports the prediction that recall is biased toward the mean. The slope for each category was significantly smaller than zero (female, *t* (21) = -6.775, *p* <.001; male, *t* (24) = -6.678, *p* < .001; masked, *t* (24) = -10.823, *p* < .001).

The difference between intercepts tests our key prediction that people studied at the same absolute study size can be differentially biased, depending on prior knowledge about height at the gender level. An intercept difference indicates a bias to the center of the height range for each gender, such that people that are presented at a size that is small relative to that gender's height

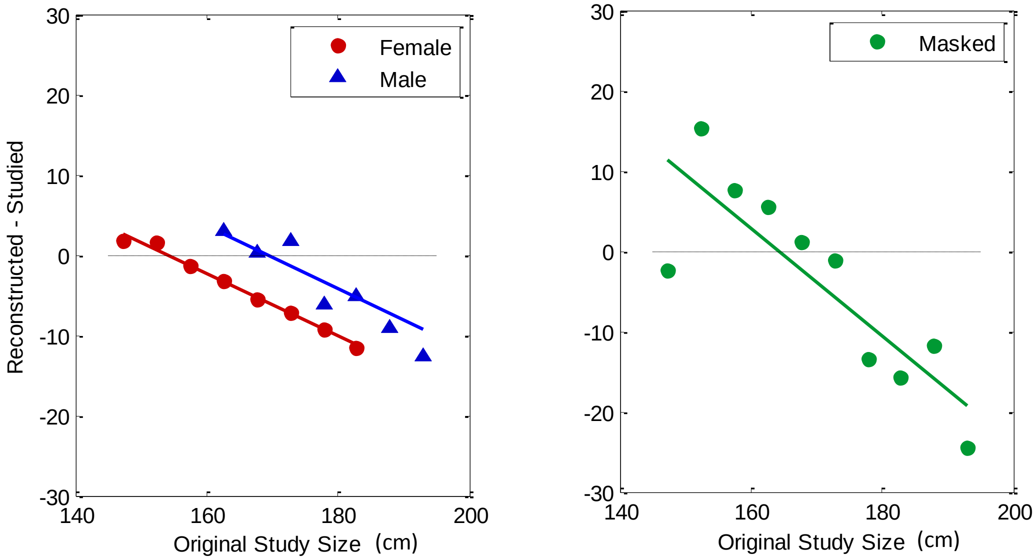


Figure 5.4. Recall bias as a function of original study size. The results are grouped by female, male and masked stimuli. Circles and triangles denote the experimental data. Lines show the linear regression model fits.

Table 5.1. *Mean slopes and intercepts by category*

	Females		Males		Masked	
	Mean	SD	Mean	SD	Mean	SD
Slope	-0.39	0.28	-0.39	0.27	-0.67	0.30
Intercepts	51.81	4.08	56.95	3.74	110.27	6.89

distribution lead to a greater overestimation error than do people presented at a relatively large size. To assess the influence of this gender-level prior knowledge, a linear regression model was fitted separately for each subject. The regression model contained three parameters for each category: two intercept parameters, corresponding to male and female, and a single slope parameter (see Figure 4 for fits of this regression model). A separate regression model was fitted to the masked category with a single intercept parameter. Table 1 shows the mean estimated slopes and intercepts across categories.

The results show that the intercept for female was smaller than that for male. This difference in intercepts by relative study size supports the prediction of gender-level prior effects. A repeated measures ANOVA with three levels (female, male and masked) found a significant effect of category [$F(2,42)=698.56$, $p< .001$]. Bonferroni adjusted contrasts showed a significant difference between female and male intercepts [$p < .001$] and between masked and both female [$p < .001$] and male [$p < .001$] intercepts. These differences are consistent with an influence of prior knowledge at a more fine grained level of knowledge. These differences in intercepts confirm the prediction that when a male and a female are studied at the same height, reconstruction is differentially biased depending on their relative height.

Experiment 2

Having demonstrated the hierarchical effects of the influence of prior knowledge on recall for height we sought to investigate whether prior knowledge can lead to an average improvement in recall (Huttenlocher, et al., 1991; Hemmer & Steyvers, 2009a; 2009b). To this end, we implemented an experiment to measure people's prior knowledge for human height. In the experimental task, participants were asked to give subjective estimates for the height of people they have never seen before. From these estimates we can construct a psychological distribution of heights. The subjective estimates correspond to a memory task where the stimulus has not been studied before, i.e., zero study time. This will allow for a comparison between estimating heights from prior knowledge and recalling heights from memory, as well as evaluating the interaction between prior knowledge and memory for masked and unmasked figures. In other words, we can investigate whether people are better at estimating height for unmasked images of males and females even when they have not previously studied the true height.

A second goal of this study is to investigate a potential confound in Experiment 1 – people could have guessed at the true height of a person based on some configural information (e.g., the size of their head or the width of their shoulders) other than gender. The subjective height estimates obtained in this experiment allow us to measure the degree to which configural information contribute to the height estimation, i.e., to what degree can one guess the height of a person merely based on configural information?

Participants

There were 22 participants, all students at the University of California, Irvine. Participants of Experiment 2 were not involved in the Experiment 1. They were compensated with course credit.

Procedure

To assess people’s prior knowledge for the heights of females, males and ambiguous silhouettes, we asked people to make height judgments for all stimuli. The stimuli in this experiment were identical to Experiment 1. The 72 images were blocked and randomized in the same manner as in Experiment 1. The comparison image of the door that was used in Experiment 1 remained fixed on the left side of the screen at all times. The stimuli were presented sequentially on the right side of the computer screen. The height of each stimulus was initialized at a random size. Participants were asked the following question for each image: “What do you think is the true height of this person, compared to the door on the left?” The task for the participant was to estimate the true height of the stimuli without knowledge of the actual height. They responded by using a slider to resize the image of the person.

Results

All of our measurements will be given in centimeter (cm) units. To determine the relationship between episodic memory (Experiment 1) and prior knowledge (Experiment 2), we assessed recall error as mean absolute deviation. This measures the absolute deviation between the recalled and studied height for the memory data, and the absolute deviation between the estimated and true height for the prior knowledge data. Larger values indicate worse performance. Furthermore, we assessed chance performance for each category by calculating the mean absolute difference between the studied height and a random prior knowledge response from that category. This is the equivalent of the participants randomly guessing on each trial.

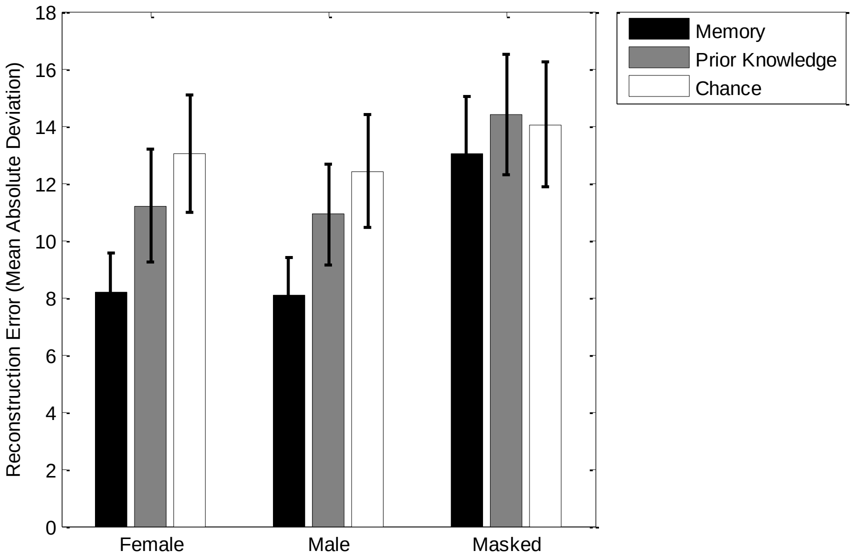


Figure 5.5. Recall error measured by mean absolute deviation. The error is grouped by category and study type. The baseline error is equivalent to responses based on the mean of the overall height distribution. Error bars indicate standard errors of the means.

This measure contains no configural information, no prior knowledge and no memory of the true height. Figure 5 shows the mean absolute deviation by category and experiment type as well as chance performance for each category. A 2 (experiment type) x 3 (category) repeated measures ANOVA was performed. There was a significant main effect of experiment type [$F(1,42) = 759.749$, $p < .001$]. Recall error was significantly lower for memory responses than for prior knowledge responses. There was a significant main effect of category [$F(2,84) = 58.164$, $p < .001$]. Bonferroni adjusted contrasts showed that recall error was greater for masked figures than for female [$p < .001$] and for male [$p < .001$]. There was no significant difference between the recall errors for female and for male [$p=1.000$]. The worst performance was observed for masked figure in the prior knowledge experiment. In this condition, performance was similar to chance, (indicated in Figure 6 by the white bar). There was no significant interaction effect between experiment type and category [$F(2,84) = 2.003$, $p = .14$]. However, a trend was

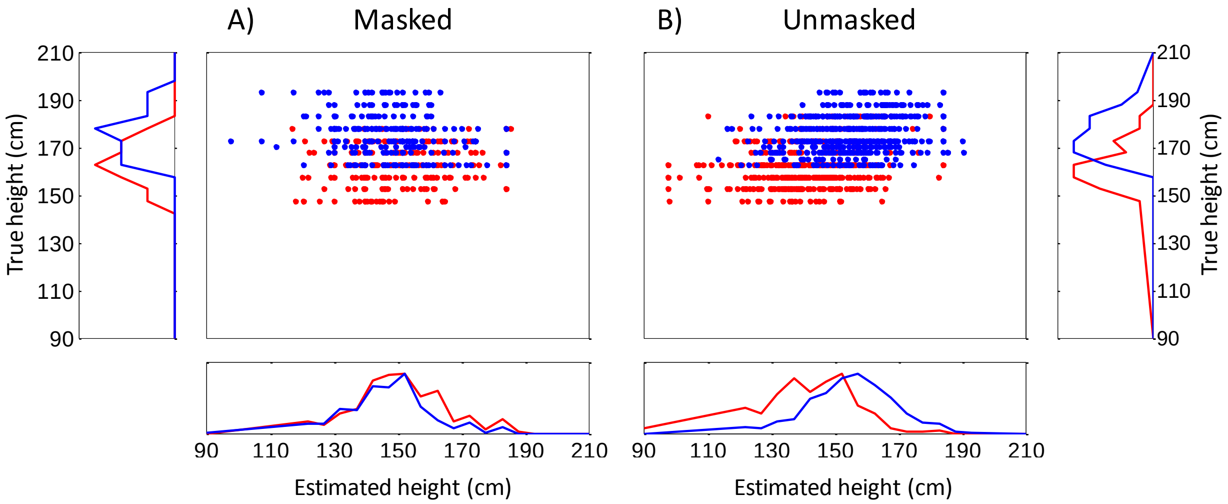


Figure5.6. Correlation between the estimated and true heights of masked female and male stimuli. Red indicates female, while blue indicates male. B) Correlation between the estimated and true heights of unmasked female and male stimuli.

observed such that in the prior knowledge condition people performed better in both the female and male category than in the masked category.

Guessing about the height of a person can be based both on gender and configural cues. In order to disentangle these effects, we analyzed the relationship between estimated and true height for the masked and unmasked stimuli, and separately for the female and male stimuli. Figure 6 panel A shows that there was no difference between the estimated height of females and males for the masked stimuli. While a clear difference exists between the true height (y-axis) of females (red points) and males (blue points), no such difference is observed for estimated height (x-axis)[3]. This shows that we effectively removed all gender cues from the masked stimuli. More interestingly, Figure 6 panel B shows the effect of both gender and configural cues. First, there is

---

[3] We checked to ensure that the female figures in the masked images were not systematically judged to be shorter than the male masked images. A paired t-test showed that the estimated means for females (151.04 cm) were not smaller than the estimated means for males (147.24 cm), $t(526)=3.12$, $p=0.99$. The female figures were instead judged to be taller than the male figures on average. This might have been a result of our effort to remove female gender characteristics. This did not affect the results of the memory study, however.

a clear separation between females and males for both the true and the estimated height (see the marginal distributions in the bottom and side panel). These differences in the data patterns between the unmasked and masked stimuli can be explained by the discrepancy in availability of gender cues. The unmasked stimuli provides gender cues which might allow participants to use their knowledge of height and gender – namely, that females are shorter than males – in estimating the true height. This is demonstrated by the close correspondence between the a priori expectations people have for the heights of males and females and the true distribution of heights in the population. The masked stimuli do not provide a cue for gender and thus should not allow the participants to disambiguate the heights of females and males. Second, within each gender population in the unmasked stimuli, there is a small but significant correlation between the estimated and true height – females (r=0.24) and for males (r=0.19). In other words, just looking at a particular female or male image, without knowing their true height, it is possible to predict their true height. This suggests that people might be able to use information other than gender (i.e., configural information) to estimate the true height of a person. The fact that there are configural effects present a potential confound in interpreting the results of Experiment 1. This is because memory performance might potentially have been guided by configural effects in addition to prior knowledge.

To examine the strength of the configural effect, we compared chance performance to each experiment type. The chance measure (described above) contains no configural information, no prior knowledge and no memory of the true height. The difference between chance and prior knowledge is a measure of the influence of configural information as well as the impact of prior knowledge. The difference between prior knowledge and recall is a measure of the influence of memory. Figure 5 shows the mean absolute deviation for prior knowledge (gray bars), memory (black bars), as well as chance (white bars). While there is a reduction in error form the chance measure to prior knowledge performance in both the female and male category, there is also a further reduction in error from prior knowledge to memory performance. This means that while configural information has an impact on performance, configural information alone cannot account for the findings in the memory experiment. This suggests that a priori, people can give an estimate of the true height of a person based only on the configural properties of the person. However, the accuracy of such an estimate is not very high and importantly, this factor by itself cannot explain the performance differences between the memory and prior knowledge experiment. Clearly, participants are relying on some memory of the individual in their height estimate.

General Discussion

In the current study we investigated the influence of prior knowledge of gender differences in height on recall for height of specific individuals. We demonstrated that episodic memory is influenced by general knowledge about height such that short people are overestimated and tall people are underestimated. Furthermore, we showed that the influence of prior knowledge is hierarchical, such that more fine-grained information about the height based on gender leads to biases toward the average height of men and women independent of the absolute study size. Recalled height for a relatively short male (shorter than the mean height for men) will be overestimated while recalled height for a relatively tall female (taller than the mean height for women) will be underestimated. Overall, our results are consistent with the view that prior knowledge influences recall at multiple levels of abstraction.

It is possible that standard effects such as sequential effect, slider effects and edge effects might have acted on our data. However, our experimental design precludes the possibility that they are the sole explanation of our findings – see Hemmer and Steyvers (2009a) for a detailed account. There are three possible ways that people might reconstruct the studied height: 1) they might recall the height from memory; 2) they might use prior knowledge to reconstruct the height of the person; 3) they might guess at the height using some artifact information inherent in the stimulus, such as configural features, that might indicate if a person is short or tall. Our analysis of the prior expectations for female and male height showed that people might be able to guess the true height of a person based on some configural information. This alone cannot explain our data because of the difference in performance between the memory and prior knowledge experiments. Furthermore, while gender information was removed from the masked stimuli, other configural information was not completely removed along with it. Thus, if height estimation was to be influenced by configural information this should have been the case for both the masked and unmasked stimuli.

We have also shown that having gender-level knowledge improves recall performance compared to ambiguous stimuli with a broader level of prior knowledge. While having episodic information leads to the best performance, there is less error in size estimation for gendered stimuli than for ambiguous stimuli. When making height estimates about ambiguous stimuli, the best strategy a participant could employ is to guess with the overall average height of people. However, when gender information is available, participants can guess with knowledge about the height at a more fine-grained level of gender. The results obtained from our experiment suggest that prior knowledge can be beneficial towards the accuracy of our memory.

As our findings are applied to the scenario discussed at the beginning of this paper, knowing the gender of the crime suspect prior to making a height approximation would help yield a more accurate estimate. Using prior knowledge can be an efficient strategy to reconstruct episodic memories, especially if those memories are noisy and incomplete.

Chapter 6

Summary and Conclusions

In the work reviewed in the previous chapters I investigated the interactions of episodic and semantic components in reconstructive memory across multiple task domains. The goal of these studies has been to address the same underlying questions: What prior knowledge do participants bring to the task? What is the relative contribution of prior knowledge to recall? How do episodic and semantic components interact in recall?

Through the presented studies I have worked to characterize people's knowledge of the environment and have demonstrated the people have a strong representation of their environment. This knowledge in turn influences how people perceive and recall events related to their environment. While the assertions of more traditional approaches in memory research, such as false memory, are that prior knowledge leads to errors in recall, the findings presented here suggest that this is not the complete story. When reconstructive memory is tested using naturalistic and ecologically valid stimuli, that are representative of the environment, and for which people have prior expectations, several interesting and noteworthy findings result.

First, I found that prior knowledge interacts with episodic memory at multiple levels of abstraction and the combination of prior knowledge and episodic memory is dependent on familiarity. Second, 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 consistent with previous findings (Konkle & Oliva, 2007; Kan, Alexander, Verfaelle, 2009; Minsky, 1975; Huttenlocher et al. 1991). Third, base line performance from guessing with prior knowledge is quite high. This is particularly interesting, because one would assume episodic memory to be the main source for reconstructing events. While it is expected that prior knowledge can be used to fill in gaps in memory it was unexpected that the semantic contribution was so large. Fourth, items with high levels of prior knowledge are associated with both a higher rate of recall and a lower intrusion rate than items with low levels of prior knowledge. Objects with no prior expectation (i.e., incongruent objects) are also recalled at a higher rate than item with low prior expectation. Brewer and Treyens (1981) showed that objects consistent with many scenes are better remembered. Pezdek et al. (1989) showed that incongruent objects are better remembered. However, when using naturalistic stimuli both effects might occur simultaneously.

The interaction between episodic memory and prior knowledge is not an isolated phenomenon. These findings are robust and the interactions occur across multiple domains. Together these studies provide converging evidence of the quality of prior knowledge and the influences of prior knowledge on reconstructive memory. Using ecologically valid situations where prior knowledge is in alignment with the stimulus material to be remembered, has demonstrated that prior knowledge plays a valuable role in memory retrieval and can actually increase the accuracy of our memories.

References

Altmann, E. M. (2003) Reconstructing the serial order of events: A case study of September 11, 2001. *Applied Cognitive Psychology*, **17**, 1067-1080.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

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

Biernat, M. (1993). Gender and height: Developmental patterns in knowledge and use of an accurate stereotype. *Sex Roles*, **29**, 691- 713.

Bower, G. H., Karlin, M. B., & Dueck, A. (1975). Comprehension and memory for pictures. *Memory & Cognition*, **3**, 216-220.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory. *Cognitive Psychology*, **11**, 177-220.

Brainerd, C. J., Wright, R., & Reyna, V. F. (2002). Dual-retrieval processes in free and associative recall. *Journal of Memory and Language*, **46**, 120–152.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior,* **11**, 717-726.

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

Budescu, D. V. & Yu, H. (2007). Aggregation of opinions based on correlated cues and advisors. Journal of Behavioral Decision Making, **20**, 153-177.

Carmichael, L. C., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, **15**, 73-86.

Crawford, E., Huttenlocher, J., & Engebretson, P. H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, **11**, 280-284.

Eagly, A. H, & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, **46**(4), 735-754.

Estes, W.K. (1997). Processes of Memory Loss, Recovery, and Distortion. Psychological Review, **104**, 148-169.

Fahsing, I.A., Ask, K., & Granhag, P.A. (2004). The man behind the mask: Accuracy and predictors of eyewitness offender descriptions. *Journal of Applied Psychology*, **89**, 722-729.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology: General, **108**, 316-355.

Gagon, L.M. & Dixon, R.A. (2008). Remembering and retelling stories in individual and collaborative contexts. *Applied Cognitive Psychology*, **22**, 1275-1297.

Galton, F. (1907). Vox Populi. *Nature*, **75**, 450-451.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2003). *Bayesian Data Analysis*. Chapman & Hall.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, **114**, 211-244.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Science*, **101**, 5228-5235.

Healy, A. F., Havas, D. A., Parker, J. T. (2000). Comparing Serial Position Effects in Semantic and Episodic Memory Using Reconstruction of Order Tasks. *Journal of memory and language*, **42**, 147-167.

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review,* **16**, *80-87.*

Hemmer, P. & Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. *Topics in Cognitive Science,* **1**, 189-202.

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* (pp. 1557-1562). Austin, TX: Cognitive Science Society.

Hemmer, P. & Steyvers, M. (2008). A Bayesian Account of Reconstructive Memory. In V. Sloutsky, B. Love, and K. McRae (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum

Hogarth, R. M. (1978). A note on aggregating opinions. Organizational Behavior and Human Performance, **21**(1), 40-46.

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.

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review,* **149**, 288-294.

Kan, I.P., Alexander, M.P. & Verfaellie, M. (2009). 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 & Psychophysics*, *60*, 1318-1328.

Kerst, S. M., & Howard, J. H., Jr. (1978). Memory psychophysics for visual area and length. *Memory & Cognition*, **6**, 327-335.

Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in perception and memory. In D. S. McNamara & J. G. Trafton (Eds.), *Proc.s 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.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory. *Annual Review of Psychology*, **51**, 481-537.

Loftus, E.F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology,* **7**, 560-572.

Marden, J. I. (1995). Analyzing and Modeling Rank Data. New York, NY: Chapman & Hall USA.

McDowell, M., Fryar, C.D., Ogden, C.L. & Flegal, K.M. (2008). Anthropometric Reference Data for Children and Adults: United States, 2003-2006. *National Health Statistics Reports*, *10*.

Miller, B. Hemmer, P. Steyvers, M. & Lee, M. D. (2009). The wisdom of crowds in ordering aggregation problems. *Proceedings of the 9th International Conference of Cognitive Modeling.*

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Moyer, R. S., Bradley, D. R., Sorensen, M. H., Whiting, C., & Mansfield, D. P. (1978). Psychophysical functions for perceived and remembered size. *Science*, **200**, 330-332.

Nairne, J. S. (1992). The loss of positional certainty in long-term memory. *Psychological Science*, **3**, 199–202.

Nelson, T. E., Biernat, M. R., & Manis, M. (1990). Everyday base rates (sex stereotypes): Potent and resilient. *Journal of Personality and Social Psychology*, **59**(4), 664-675.

Otgaar, H., Candel, I., Scoboria, A., & Merkelbach, H. (2010). Script knowledge enhances the development of children's false memories. *Acta Psycgologica*, **133**, 57-63

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.

Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory*, **14**, 289-296.

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

Rothbart, M., Fulero, S., Jensen, C., Howard, J. & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology,* **14**, 237-255.

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

Sa, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, **91**(3), 497-510.

Sailor, K. M., & Antoine, M. (2005). Is memory for stimulus magnitude Bayesian? *Memory & Cognition*, **33**, 840-851.

Stasser, G., Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of personality and social psychology*, **48**(6), 1467-1478.

Steyvers, M., & Griffiths, T. (2008). Rational analysis as a link between human and information retrieval. In: N. carter and M. Oaksford (eds.) *The Probabilistic Mind: Prospects from Rational Models of Cognition.* Oxford University Press.

Steyvers, M., Lee, M.D., Miller, B., & Hemmer, P. (2009). The Wisdom of Crowds in the Recollection of Order Information. In J. Lafferty, C. Williams (Eds.) *Advances in Neural Information Processing Systems, 23*. MIT Press.

Surowiecki, J. (2004*). The Wisdom of Crowds*. New York, NY: W. W. Norton & Company, Inc.

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.

Wallsten, T.S., Budescu, D.V., Erev, I. & Diederich, A. (1997). Evaluating and combining subjective probability estimates. Journal of Behavioral Decision Making, 10, 243-268.

Yi, S. K. M., Steyvers, M., Lee, M. D., Dry, M. J. (2010) Wisdom of the Crowds in Minimum Spanning Tree Problems. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.