THE UNIVERSITY OF CALGARY

It's in the Details: Effects of Test-List Context on Memory for an Event

by

Denise D. L. Richardson

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF PSYCHOLOGY

CALGARY, ALBERTA

SEPTEMBER, 2005

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled " It's in the Details: Effects of Test-List Context on Memory for an Event", submitted by Denise D. L. Richardson in partial fulfillment of the requirement for the degree of Master of Science.

Supervisor, Dr. G. Bodner, Department of Psychology

Dr. C. Sears, Department of Psychology

Dr. J. Ellard, Department of Psychology

Dr. J. Mueller, Division of Applied Psychology

Sept 8/05

Date

Abstract

This study examined how recognition judgments for a set of event details are influenced by the other details on the test. Participants viewed a movie clip of a crime scene, then assigned remember/know/neither judgments to details on a recognition test. Easy, medium, and hard details, based on probability of recognition, were established in Experiment 1. In Experiment 2, medium details received more remember judgments when they were mixed with hard details relative to when they were mixed with easy details. Similarly, in Experiment 3, a block of medium details received more remember judgments when preceded by a block of hard details than when preceded by a block of easy details. Informing participants of the difficulty of each block of details eliminated these test-list context effects in Experiment 4. The test-list context influenced participants' functional definitions of remembering, which in turn affected how they classified their memories of the event details.

Acknowledgements

I would like to thank Andreas Breuer for his programming expertise that made these experiments possible, and my committee members Dr. Chris Sears, Dr. John Ellard, and Dr. John Mueller for their time and energy during my defense. I would also like to acknowledge Dr. Glen Bodner's guidance and inspiration from his past research that provided the bases for this project. Finally, I want to thank Dr. Michelle Eskritt for initiating my interest in cognition in the first place.

Dedication

I dedicate this thesis to the four people that I love most in this world. Jessica, you are my dear sweet daughter, whose presence continually inspires me to strive to be the best that I can be. Michael, your sense of justice that inspires me to be kind and just to everyone I come in contact with. Matthew, who always reminds me of how important it is to enjoy life and to love with all of my heart. Finally, to my dear sweet Ross, who serves as the foundation and icing of my daily existence. I love each of you with all of my heart, and I look forward to creating many memories with each of you.

TABLE OF CONTENTS

List of Tables

List of Figures

It's in the Details: Effects of Test-List Context on Memory for an Event

After viewing a crime event, witnesses are often asked to report what they saw. Their memory reports have implications for crime scene investigations and courtroom testimony. That is, witnesses' reports of what they remember can influence whether certain aspects of a crime investigation are followed up, and can also influence whether a suspect is eventually found guilty in the courtroom. Because eyewitness memory reports have important implications, researchers have examined the factors that affect the way witnesses report their memory for an event. Some researchers have focused on whether testing memory in the original environment where learning took place enhances memory accuracy (e.g., Brown, 2003; Smith, 1988). Other researchers have focused on aspects of the memory test itself, and examined the influence of question/item difficulty on memory accuracy (e.g., Marquis, Marshall, Oskamp, 1972; Schraw & Roedel, 1994).

The present study examined how memory for the details of a witnessed event is influenced by the mixture and order of the other details presented during the recognition test/interview. This is an important issue because investigators do not have control over how events are encoded, but they do have control over the way witnesses' memory is examined, which can affect what witnesses report (e.g., Loftus, 1992). For example, Loftus and Palmer (1974) found that participants asked "How fast were the cars going when they *smashed* into each other?" reported higher speed estimates than those asked "How fast were the cars going when they *contacted* each other?", showing that test wording can influence eyewitness testimony.

Sometimes the format of a memory test can even cause people to report remembering event details that did not occur. For example, presenting misinformation

about a witnessed event after the fact has been found to interfere with memory for the original event, resulting in people mistakenly reporting the new information as part of the original event (i.e., the misinformation effect). McCloskey and Zaragoza (1985) found that the options provided on a 2AFC (alternative-forced choice) recognition test influenced whether people reported having witnessed the misinformation or the original event detail. Studies such as Loftus and Palmer (1974) and McCloskey and Zaragoza (1985) highlight the importance of examining the influence of test-based factors on eyewitness memory reports.

Effects of Test Format on Memory for an Event

The effects of the mixture and order of items on a test of memory for an event have not been studied, but other aspects of test format (question type, item difficulty) have been studied. Marquis et al. (1972) had participants view a film clip and then tested their memory accuracy with various question types (free recall, leading questions, multiple choice). Question type had a large impact on memory for hard items, but had little impact on memory for easy items. After witnessing an event, Ibabe and Sporer (2004) gave participants received one of three question types: open-ended, 4AFC (i.e., multiple choice), or true/false. Responses to open-ended and true/false questions were more accurate than responses to 4AFC questions. The similar accuracy for open-ended and true/false questions revealed that structured questions do not always result in lower accuracy as was previously thought (e.g., Marsten, 1924; Stern, 1910), suggesting that structured questions can be appropriate for testing memory for an event.

Item Difficulty

Influence of item difficulty on subjective memory reports. Winkielman and Schwartz (2001) reported an influence of easy versus difficult test contexts on subjective memory reports for childhood events. Participants recalled either four (easy recall condition) or 12 (difficult recall condition) events from their childhood, and were then told either that a happy or unhappy childhood is difficult to remember (belief manipulation). Finally, they completed a survey that assessed whether they believed they had a happy or unhappy childhood. Participants in the difficult recall condition reported a happier childhood when they were told that pleasant events are difficult to recall, and reported a less happy childhood when they were told that unpleasant events were more difficult to recall, whereas ratings of childhood happiness were not affected by the belief manipulation in the easy recall condition.

Winkielman and Schwartz argued that participants in the difficult recall condition were likely surprised at the difficulty of recalling 12 childhood events and hence attributed their unexpected difficulty to the pleasantness or unpleasantness of their childhood. In contrast, participants in the easy recall condition did not experience a feeling of surprise because they were able to recall the four childhood events, and hence they were not influenced by the belief manipulation. Thus, the difficulty of recalling childhood events influenced the way people reconstructed the happiness of their childhood.

Accuracy-confidence, and item difficulty in an eyewitness paradigm. Other studies have used recognition tasks to examine how item difficulty affects the correlation between accuracy and confidence. For example, Kebbell, Wagstaff, and Covey's (1996,

Experiment 1) participants viewed a six-minute movie clip of an accident and then answered 20 2AFC questions. Ten of the 20 questions were considered easy (e.g., "What sex was the person in the bed?"), and 10 were considered hard (e.g., "Was there a picture of a zebra or horse on the wall?"). Participants also rated their confidence in each response. It was found that easy questions were answered more accurately and confidently than hard questions, resulting in a very strong accuracy-confidence relationship for easy questions, and a weak accuracy-confidence relationship for hard questions (see also Schraw & Roedel, 1994). To further examine the idea that differences in test item difficulty affect the accuracy-confidence relationship, in Experiment 2 open-ended questions were divided into three different categories (easy, medium, hard). Kebbel et al. found that the relationship between confidence and accuracy became considerably weaker as question difficulty increased. These results raise the question of how between-group manipulations of the test context, based on item difficulty, might influence witnesses' subjective memory reports. The present set of experiments examined how reports about subjective memory experiences within the remember/know paradigm were influenced by the test-list context.

Remember/Know Recognition Judgments

Remember/know judgments are often used to evaluate the subjective, phenomenological experience that accompanies an experience of recognition. People appear to be able to distinguish between two subjectively distinct types of recognition experiences. *Remembering* occurs when recognition is accompanied by recollection of episodic details (e.g., thoughts, feelings, images that occurred at the time of encoding), whereas *knowing* occurs when a stimulus seems familiar but the person is unable to

recollect any episodic details about his/her earlier experience(s) with it (Tulving, 1985). Fisher, Geiselman, and Amador (1989) suggested that asking witnesses to provide remember/know judgments helps to fine-tune the interview procedure and results in less false information being reported. Remember/know judgments focus participants on the subjective and phenomenal aspects of their recognition experiences, whereas simple yes/no judgments do not.

Several accounts of what the remember/know distinction may be tapping into have been posited. Tulving (1985) suggested that the phenomenal experiences of remembering and of knowing may be produced by activation within different memory systems. Specifically, the *memory systems account* suggests that activation within the episodic memory system results in remember experiences, whereas activation within a semantic and/or perceptual memory system results in know experiences (Gardiner & Java, 1993). According to a *dual-process account*, remember judgments are the result of successful use of a recollection process, whereas know judgments reflect use of a familiarity-based process (Joordens & Hockley, 2000; Yonelinas, 2002). Another processing-based account is Rajaram's (1993, 1996, 1998; Rajaram & Geraci, 2000) *distinctiveness-fluency account*, which posits that remember experiences are driven by manipulations that increase processing distinctiveness, whereas know experiences are driven by manipulations that increase processing fluency.

Another account is based on signal detection theory. The *signal detection account* posits that recognition judgments are based on an underlying dimension of memory strength. Recognition experiences for old items form a distribution that has a higher mean strength than the distribution of recognition experiences formed for new items. To make a

recognition decision, people set a criterion along this strength dimension, such that items whose memory strength falls below the criterion are rejected as "new", whereas those above the criterion are endorsed as "old". To account for remember/know judgments, a second, more stringent criterion could be set, above the yes/no criterion, which participants use to distinguish remember from know experiences/judgments. Thus, by the signal detection account, remember and know reflect quantitative differences in memory strength rather than qualitative differences in phenomenal experience. Of course, one criticism of this account is that it is unclear why different placements of a criterion would produce subjective states of awareness that seem so qualitatively different.

Finally, according to Gruppuso, Lindsay, and Kelley's (1997) *functional account*, participants define remember experiences as those that are detailed enough to permit successful performance of the task at hand (e.g., identifying a person in an unusual context), whereas know experiences are those that are deemed to lack sufficient detail although they do produce a feeling of familiarity. The functional account further stipulates that the same memory information can lead to assignment of remember judgments in one situation or task and to assignment of know judgments in another situation or task. For example, imagine that while shopping, Mary runs into a man she has previously encountered, and she can only recollect that she has seen him at the shopping mall before. If asked whether she *recognizes the man as her butcher*, she may report simply "knowing" the man because she cannot recollect him as her butcher. In contrast, if simply asked whether she *recognizes the man*, she may report "remembering" him, because recollecting that she saw him in the mall is sufficient for claiming to

remember him given the question posed. Thus, the same recollective experience can be classified differently depending on the framing of the recognition question.

Test-List Context Manipulations With Word Lists

Bodner and Lindsay (2003) examined the effects of the test format on recognition judgments for word lists to test whether participants' functional definitions of remember and know would be affected. More specifically, they examined if assignment of remember/know judgments for a common set of studied items would depend on the memorability of the other studied items on the test list. Bodner and Lindsay (2003, Experiment 1) manipulated the level of processing (LOP) performed on each of two word lists presented during encoding. Participants in both the medium-with-shallow and medium-with-deep groups studied one list of words in a medium LOP task (deciding if the word was common or not) and studied the other list of words in either a shallow LOP task (deciding if the word contained the letter "A" or not) or a deep LOP task (deciding if the word be of use if stranded on a desert island or not). At test, participants assigned remember/know/neither judgments to medium LOP words that were presented among shallow LOP or deep LOP words (plus a set of nonstudied words).

The memorability of the test list-context (shallow vs. deep LOP words) did not affect the overall level of recognition (remember + know judgments) for the medium items that were common to both groups' test list. However, the medium-with-shallow group was more likely than the medium-with-deep group to assign remember judgments (and was less likely to assign know judgments) to the medium items. Bodner and Lindsay suggested that the different test-list contexts led the two groups to develop different functional definitions of remembering). To provide direct evidence for this claim,

participants in their fourth experiment were asked to provide their "strongest recollection" for each remember judgment they made. Consistent with the two groups having formed different functional definitions of remembering/knowing as a function of their experiences during the test, the relative weighting of particular aspects of their recollective experience differed between the two groups.

The groups' differential reliance on qualitatively different aspects of their experience when assigning remember judgments providing support for the functional account of recognition memory put forth by Gruppuso et al. (1997). The "strongest recollection" results were not predicted by the signal detection account, which cannot explain why qualitatively different bases were emphasized by the two groups when making remember judgments; these differences are not expected if remembering and knowing are merely quantitatively different. The memory systems account fails to explain why medium LOP words would activate the episodic memory system in one test-list context, but would active a semantic and/or perceptual memory system in another test-list context. Similarly, the dual-process account does not explain why one test-list context would result in use of a recollection process and another test-list context would result in use of a familiarity based process. Finally, the distinctiveness-fluency account does not reveal why one test-list context would increase processing distinctiveness while another would increase processing fluency for the same medium details.

The Present Experiments

Bodner and Lindsay (2003) stated that "We believe our results also have an important practical implication, namely that in any multi-question interview situation, the likelihood that people will claim to remember something from their past may depend, in

part, on how well they remember the answers to the other questions posed in the interview." (p. 578). The present set of experiments provided an analogue of Bodner and Lindsay's (2003) work by examining whether the test-list context would influence the way witnesses report their memory for a crime event.

Participants were first shown a DVD movie chapter of a crime scene as the event, and then assigned remember, know, or neither judgments to some details from the event on a recognition test. A set of false details not presented in the event were also included on the test. Contrary to Bodner and Lindsay (2003), the recognizability of the details on the test-list was varied, rather than also manipulating how the details were processed at encoding (i.e., LOP). That is, all participants received the same event at study, but the recognizability of the details on the test was manipulated across groups. To this end, the recognition probability for a large number of details from the event was determined through pilot testing. Details with very low recognition rates were classified as hard details, those with average recognition rates were classified as medium details, and those with very high recognition rates were classified as easy details. The details selected from pilot testing were re-tested in Experiment 1 to ensure they were categorized appropriately (e.g., easy, medium, hard), and that the categories of details were significantly different from one another.

A set of false details selected from the pilot testing were also re-tested in Experiment 1 to ensure that that they were plausible. The assignment of true/false details could not be counterbalanced given that the event was a movie chapter. Therefore, choosing plausible false details that participants would sometimes endorse was essential, to allow possible differences in response bias (i.e., willingness to claim to recognize a

detail) between the two groups to be detected. Differences in response bias would not be detectable if the false details were not potent enough to elicit recognition judgments. The careful verification of the event details in Experiment 1 ensured that the test-list context manipulations described below, and used in Experiments 2-4, were as potent as possible.

Experiment 1: Verifying Detail Selection

The primary purpose of Experiment 1 was to establish sets of easy, medium, hard, and false details, based on probability of recognition, for the test-list context manipulations implemented in Experiments 2-4. The second purpose of Experiment 1 was to ensure that participants found the event realistic and compelling.

Method

Participants. University of Calgary undergraduates participated in these experiments in return for extra course credit. No participant was tested in more than one experiment. There were 20 participants in Experiment 1.

Materials. The event was a complete 6-minute chapter titled "Dirty Cop" from the movie *Water's Edge* (Kahn, 2004, ch.5). The event begins with a heated argument between a husband and wife in a cottage. During the argument the husband takes a shotgun and runs into the woods, where he spots a police car and hides behind a tree. A police officer emerges from the car and pulls a beaten, handcuffed woman out of the back seat. The officer berates the victim before putting his handgun to her head, at which time the husband makes a noise. Worried that he is not alone, the officer throws his handgun into the car, and prepares to hit the victim over the head with a large rock. The husband emerges from behind the tree and threatens to shoot the officer. During the standoff, the officer reaches for his handgun, and is shot and killed by the husband. The husband drops

the shotgun, checks the officer's pulse, and carries the victim to safety. This chapter was chosen because: (1) it is tense and compelling; (2) it follows one character (the husband); (3) it is self-contained with a beginning, a buildup of tension, a violent climax, and a resolution; (4) it was complex enough to provide a sufficient pool of easy, medium, and hard details; and (5) almost no participants had viewed it.

For the recognition test, participants were presented with various types of observable details from the chapter (e.g., perceptual, conversational), chosen based on their recognition probability in a pilot study not described here. The set of details consisted of 15 easy details with recognition rates of .80 to .95, 15 hard details with recognition rates of .05 to .15, 15 medium details with recognition rates closest to .50 (the range was .40 to .70), and 15 plausible and potent false details with recognition rates of .10 to .25. The easy, medium, and hard detail sets contained similar numbers of emotional details, and of details from the beginning, middle, and end segments of the chapter, and contained details of similar length. More of the hard details emphasized perceptual details, relative to the easy and medium detail sets. The presentation order of the details on the recognition test was freshly randomized for each participant. Table 1 presents two sample details from each of the four conditions.

Procedure. Participants were tested individually after consenting to view and make judgments about a graphic movie clip. They were not told that their memory for the event would be tested; rather, they were told that the purpose of the study was to examine how they would judge various aspects of an emotional movie clip. Each person sat in front of a computer monitor, and the interviewer sat across from him/her in front of a second monitor, which was connected to the same computer. After watching the event on

Table 1

Sample Easy, Medium, Hard, and False Details

Detail Type	Sample Details
Easy	The husband had short brown hair.
	The officer warned the victim not to move.
Medium	The husband's name was Robert.
	The victim had a scrape on her cheek.
Hard	The husband said the typewriter was faulty.
	The number on the police car was 023.
False	The car had POLICE written on it.
	The wife said the gun was for protection.

a laptop computer, a short questionnaire was presented on the computer screen. The questionnaire examined participants' reactions to various aspects of the movie chapter (see Table 2), and determined whether they had seen the movie before. In addition to providing a manipulation check, the questionnaire served as a distracter task that allowed any emotional reactions to the event to dissipate. Participants keyed in responses with the keyboard and were assured that the interviewer could not see their responses.

The recognition test was then presented. Participants were told that details would be presented on the computer screen one at a time, and that their task was to assign one of three recognition judgments to each detail. It was explained that remember judgments

were to be assigned to details only if they were certain that the detail was presented in the movie chapter, and this recognition certainty was accompanied by episodic recollection (e.g., a thought, image, or feeling). Know judgments were to be assigned if they were certain they had viewed the detail but had no episodic recollection of it (e.g., familiarity). Neither judgments were to be assigned to all other details, including details that participants were uncertain about. It was explained that some of the details on the test were not present in the movie clip. The details were presented one at a time on the computer screen. Participants made a verbal remember/know/neither response for each detail. After the recognition test, participants were asked "How good do you think your memory typically is?", and provided answers on a seven point scale where 1 was poor and 7 was excellent. They were also asked "How difficult do you think the test was?", and provided answers from another seven point scale where 1 was very easy and 7 was very hard.

Results

Recognition judgments for witnessed details. To ensure that the easy, medium, and hard details selected from the pilot study were different from one another, separate one-way ANOVAs were performed on the mean overall recognition (sum of remember + know judgments), remember, and know judgments shown in Figure 1. The significance level was set at .05 in all experiments. Eta squared (η^2)—the proportion of variance accounted for by the independent variable—is also reported for all tests of significance where $F > 1$. The one-way ANOVAs and all pairwise follow-up tests revealed that easy, medium, and hard details were significantly different, all $ps < .001$, except know

judgments for easy and medium details were equal F (1, 19) = 1.74, MSE = .05, p = .20, η^2 = .65.

Questionnaire responses

Table 2. *Mean Judgment Ratings of the Movie Clip*

Question	
How suspenseful was the movie clip?	5.1
How compelling was the movie clip?	5.3
How graphic was the movie clip?	4.2
How absorbing was the movie clip?	5.3
How upsetting was the movie clip?	4.7
How funny was the movie clip?	1.5
How realistic was the movie clip?	3.8

Note: Judgments were made on a 7-pt scale (1 = not at all, 4 = somewhat, 7 = extremely).

Recognition judgments for false details. Recognition, remember, and know rates for false details were .22, .10, and .12, respectively, suggesting that they were plausible enough to elicit recognition experiences, and even remember judgments. Indeed, the false details were more likely to be recognized (.22 vs. .11), F (1, 19) = 6.47, MSE = .02, η^2 = .87, and to receive remember judgments (.10 vs. .02), F (1, 19) = 14.55, MSE = .001, η^2 = .94, than the hard details.

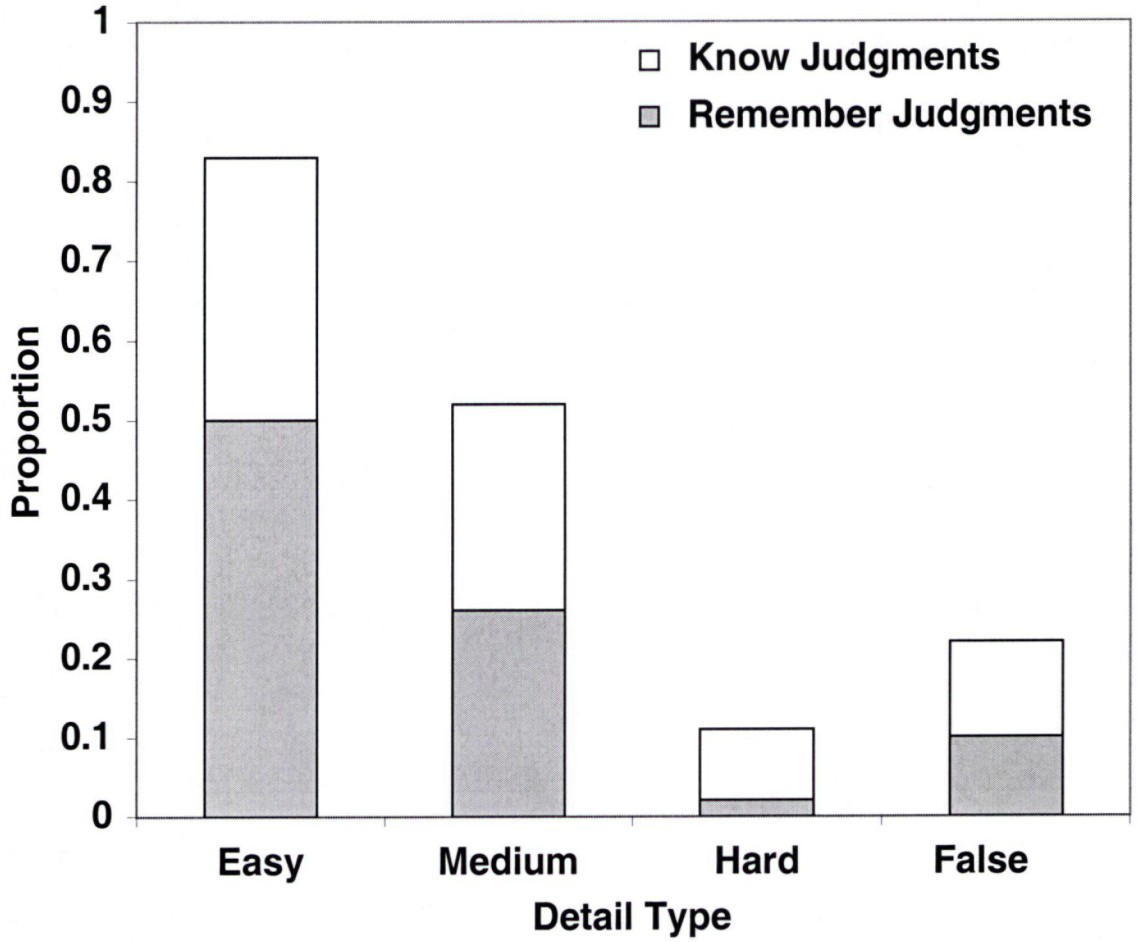

Figure 1. Mean proportion of remember and know judgments for easy, medium, hard, and false details in Experiment 1. The sum of remember and know judgments estimates overall recognition.

Discussion

Experiment 1 verified that easy, medium, and hard details were all significantly different in their probability of being recognized and remembered. It is important to note that probability of recognizing a detail cannot be equated with the memorability of the detail in the event itself, because the wording of the detail likely also influenced its recognizability. This experiment also verified that the false details were potent enough to produce recognition experiences and false remember judgments, which is necessary to

allow for potential differences in response bias across test-list context groups to be detected in the subsequent experiments. Finally, prior studies have typically not ensured that the crime event evokes an emotional response, as would be the case in a real crime situation. Participants in Experiment 1 reported finding the event reasonably engaging, if not especially realistic. Some participants noted that it was not realistic that a police officer would perpetrate a crime, for example, which may explain why the rating of realism was not high. Overall though, the ratings suggest that the crime analogue was effective in capturing participants' attention and was taken seriously.

Experiment 2: Mixed Test-List Context Effect

The details that selected in Experiment 1 were used in Experiment 2 to investigate whether memory for event details would be influenced by different test-list contexts, as predicted by the functional account. Such a result would show that Bodner and Lindsay's (2003) results with benign word lists could be extrapolated to memory for an emotionally involving event. Experiment 2 was identical to Experiment 1, except that groups were exposed to either a random mixture of hard and medium details (*hard/medium group*) or a random mixture of easy and medium details (*easy/medium group*). Both groups also had the same set of false details randomly mixed among the correct details. The functional account predicts that the hard/medium group should provide more remember judgments (and fewer know judgments) for medium details relative to the easy/medium group. Against a background of hard details, the recollection that occurs for medium details might seem sufficient to warrant remember judgments. In contrast, against a background of easy details, the same recollective experiences for medium details might

be considered too weak to justify a remember judgment, resulting in know judgments being made instead.

Method

Participants. Forty participants were randomly assigned to either the hard/medium or the easy/medium group (n = 20 per group). To ensure that participants adhered to the remember/know instructions, one participant who falsely remembered more than 30% of the false details was replaced.

Materials, design, and procedure. The materials and procedure were the same as in Experiment 1. At test, the easy/medium group received a random mixture of 15 medium, 15 easy, and 15 false details, and the hard/medium group received a random mixture of 15 medium, 15 hard, and 15 false details.

Results

Effect of test-list context on medium details. As revealed in Figure 2, medium details were 17% more likely to be recognized overall (remember + know) when mixed with hard rather than easy details (.67 vs. .50), F (1, 19) = 11.88, MSE = .02, η^2 = .39. The test-list context also influenced the likelihood that participants claimed to remember the medium details; medium details were 15% more likely to receive remember judgments when mixed with hard rather than easy details (.42 vs. .27), F (1, 19) = 7.77, MSE = .03, η^2 = .29. Assignment of know judgments to medium details was not influenced by the hard versus easy detail context (.25 vs. .23), $F < 1$.

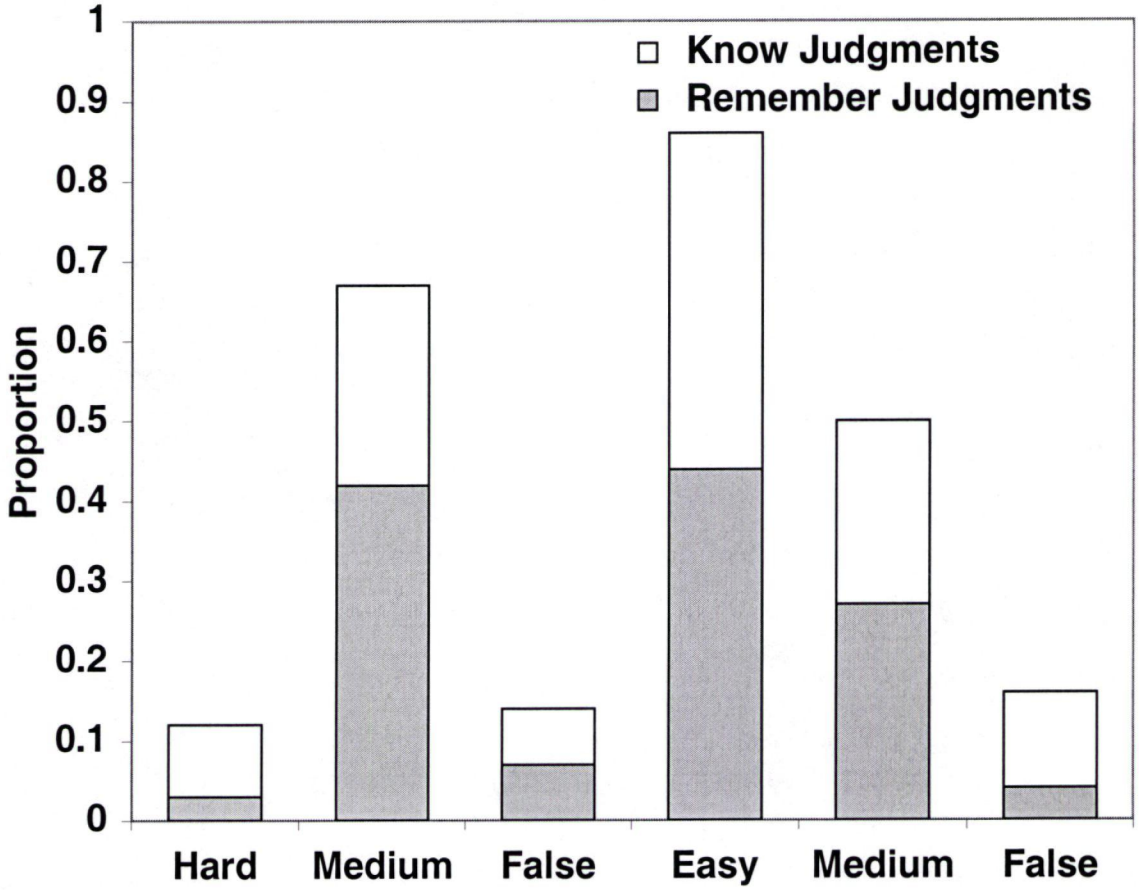

Figure 2. Mean proportion of remember and know judgments for studied and false details in the hard/medium mixture and easy/medium mixture groups in Experiment 2. The sum of remember and know judgments estimates overall recognition.

Effect of test-list context on false recognition. The hard/medium and easy/medium groups did not differ in their overall rates of false recognition (.24 vs. .16), F (1, 19) = 2.04, MSE = .05, p = .63, η^2 = .10, nor in their rates of remember judgments (.07 vs. .04), F (1, 19) = 1.17, MSE = .05, p = .59, η^2 = .07, or know judgments (.17 vs. .12), F (1, 19) = 1.05, MSE = .08, p = .68, η^2 = .09, to the false details. Because there were no significant differences in false recognition, false remember, and false know judgments between the two groups, it is unlikely that the differences in recognition judgments for medium items

were due to response bias. However, because false recognition rates differed across list contexts in Experiment 3, the effect of list context on measures of discrimination and bias are reported next, and in the remaining experiments.

Effect of test-list context on discrimination. Because the assignment of studied and nonstudied (i.e., false) details to the movie event could not be counterbalanced, the shape of their underlying strength distributions (in signal detection terms) might differ. Therefore, a simple two-high threshold model was used to compute indices of discrimination and bias, rather than a distribution-based signal-detection model. Snodgrass and Corwin (1988) found that the measures of discrimination and bias that follow from this model were independent of one another and were as appropriate as the signal-detection model was for analyzing recognition data. Discrimination refers to the participant's ability to distinguish between studied and false details, and in the two-high threshold model a discrimination score, P_r is computed by subtracting the relevant false alarm rate (the probability of endorsing a nonstudied detail) from the relevant hit rate (the probability of endorsing a studied detail). Essentially, the discrimination score attempts to "correct" the hit rate given the participant's level of response bias, and hence these are often referred to as *corrected scores.*

Discrimination scores for recognition (overall correct recognition – overall false recognition), remember judgments (correct remember – false remember), and know judgments (correct know – false know) for medium details are shown in Figure 3. These

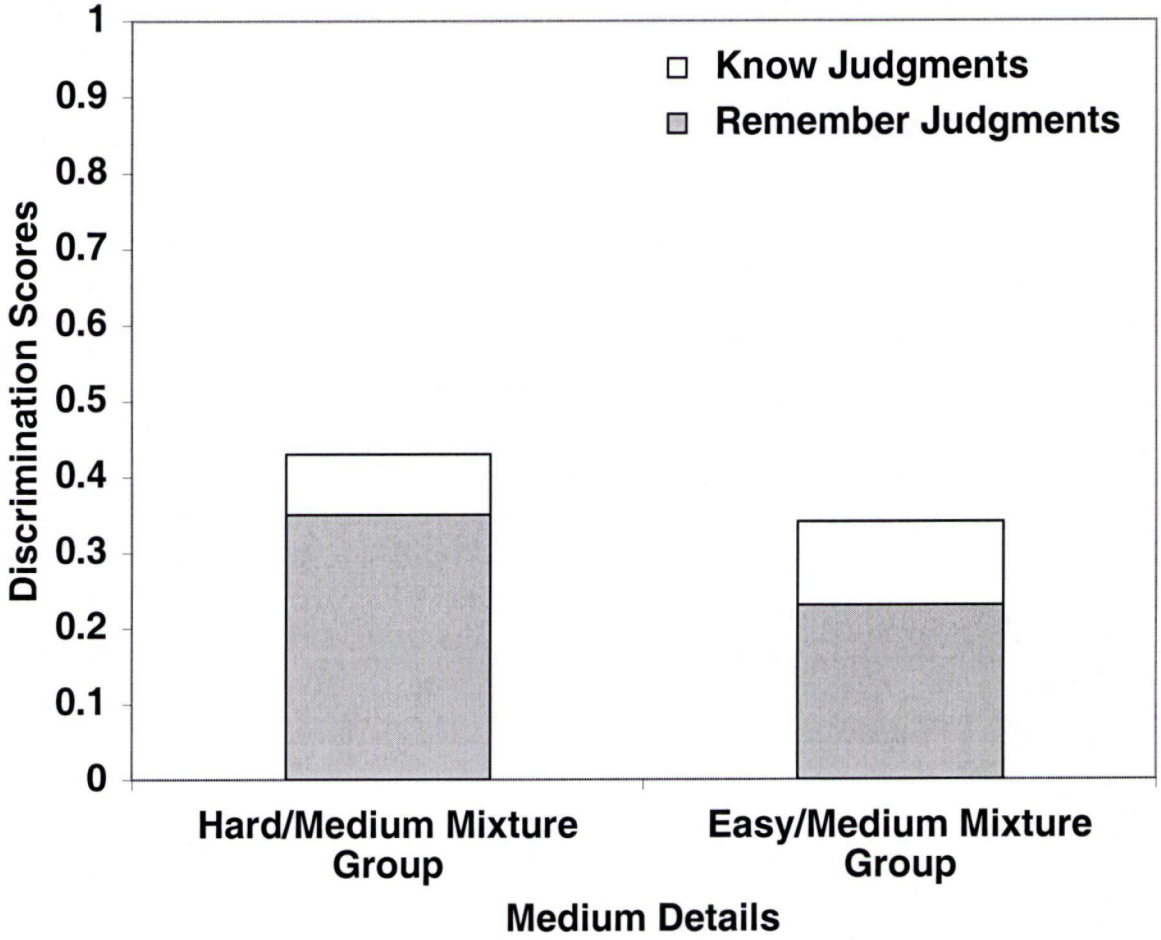

Figure 3. Mean discrimination scores for remember and know judgments for medium details in the hard/medium and easy/medium mixture groups in Experiment 2. The sum of remember and know scores estimates overall recognition discrimination.

scores were equivalent in the hard/medium and easy/medium groups for overall recognition (.44vs. .34), F (1, 19) = 1.85, MSE = .09, p = .18, η^2 = .09, and for know judgments (.08 vs. .11), $F < 1$. However, the effect of the test-list context on remember judgments remained strong; remember scores were still 12% higher in the hard/medium group than in the easy/medium group (.35 vs. .23), F (1, 19) = 4.21, MSE = .15, η^2 = .18.

Effect of test-list context on bias. Within the present experiments bias measures were computed for recognition, remember, and know judgments using the two-high

threshold bias index formula $B_r = FA/[(1-(H-FA)]$. In this formula FA indicates false-alarms and occurs when people incorrectly report recognizing a detail that was not presented, and H indicates hits, which occur when people correctly report recognizing a detail that was presented. The bias measure (B_r) determines if a participant was liberal or conservative when reporting whether he/she recognized a detail. A bias score below .5 indicates a conservative response bias, a bias score of .5 indicates no bias, and a bias score above .5 indicates a liberal response bias.

Experiment 2 bias scores for overall recognition reveal that both groups were on the conservative side, but the hard/medium group was less conservative than the easy/medium group (.38 vs. .22), F (1, 19) = 5.51, MSE = .04, η^2 = .85. However, both groups were equally conservative when making remember (.08 vs. .05), F (1, 19) = 1.58, MSE = .04, p = .36, η^2 = .29, and know judgments (.17 vs. .12), F (1, 19) = 1.64, MSE = .03, p = .27, η^2 = .53. The bias scores provide some evidence that the hard/medium group was more liberal relative to the easy/medium group in terms of willingness to say "remember" and/or "know". However, bias scores were equivalent for remember judgments, suggesting that the test-list context effects on rates of remembering were not driven by differences in response bias between the two groups.

Effect of test-list context on metamemory judgments. The hard/medium group reported finding the test more difficult than the easy/medium group (5.1 vs. 4.3), F (1, 38) = 6.08, MSE = 6.40, η^2 = .86, which was the case. More interestingly, the hard/medium group also reported having a poorer memory in general than the easy/medium group (4.5 vs. 5.2), F (1, 38) = 5.09, MSE = 4.23, η^2 = .84. The test-list context not only influenced how people felt about their test performance, it also

influenced how they felt about their memory in general. This test-list context effect on metamemory judgments fits well with Winkielman and Schwartz's (2001) finding that memory-task difficulty influenced people's opinion about their memory.

Discussion

Medium details were more likely to be classified as remembered when mixed with hard details than when mixed with easy details, extending Bodner and Lindsay's (2003) findings with word lists to memory for details from a witnessed event. The higher rates of remember judgments for the hard/medium group reflected an increase in discrimination rather than a more liberal bias. Although overall recognition (remember + know) also increased in the hard/medium group, contrary to the null effect reported by Bodner and Lindsay (2003), this difference was not present in the corrected recognition scores where the difference in bias was taken into account. The equivalent corrected recognition scores, coupled with the differences in remember scores, suggest that the test-list context affected how participants interpreted their recognition of event details, rather than affecting the amount of information they recollected.

Effects of the test-list context are not usually picked up using simple yes/no recognition judgments (Hirshman, 1995; Murnane & Shiffrin, 1991; Ratcliff, Clark, & Shiffrin, 1990; Ratcliff, Sheu, & Gronlund, 1992; Yonelinas, Hockley, & Murdock, 1992). In contrast, such effects are revealed when remember/know judgments, which focus participants on their subjective recognition experiences, are used (Bodner & Lindsay, 2003; McCabe & Balota, 2005).

Why might the two groups' classify the same recollective experience for a given medium detail differently? To take one example, being able to recollect that the

husband's name starts with "R" may have been sufficient to lead participants in the hard/medium group to assign a remember judgment to the medium detail "The husband's name was Robert". Because participants in this group were not expecting to have a rich recollective experience, due to the difficult test-list context, a weak recollection would suffice for making a remember judgment. In contrast, the easy/medium group would come to expect to have rich recollective experiences, and hence only having access to the first letter of the husband's name may have seemed more like a know experience in that context.

Experiment 3: Blocked Test-List Context Effect

Experiment 2 revealed that test-list context effects can be extrapolated beyond word lists to memory for an event. The goal of Experiment 3 was to examine whether participants' definitions of remembering are set early and maintained throughout the test (*global setting hypothesis*), or if they might be modified as a function of recent recognition experiences during the test (*local setting hypothesis*). These two possibilities were compared in Experiment 3 by presenting the details in blocks instead of mixtures. More specifically, the *hard-then-medium group* was presented with a block of hard details followed by a block of medium details, and the *easy-then-medium group* was presented with a block of easy details followed by a block of medium details. Participants were not informed about the blocked nature of the test, and the details were presented in a seamless fashion with no breaks between blocks.

If blocking eliminates the test-list context effects on medium details found in Experiment 2, this would suggest that participants change how they define remembering versus knowing when the robustness of their recognition experiences changes in the

second block. Such an outcome would imply that remembering and knowing are defined locally, on the basis of the recognition experiences on recent trials. Alternatively, if the Experiment 2 pattern is replicated, this would suggest that participants define remembering and knowing on the basis of the recognition experiences provided in the first (hard or easy) block, and maintain those definitions even when their recollective experiences undergo a dramatic shift in the second (medium) block. If functional definitions of remembering and knowing are set globally, then the presentation of only hard or only easy details during the first block should produce quite different functional definitions of remembering, and a robust test-list context effect on medium items in the second block should be obtained.

Method

Participants. Forty participants were randomly assigned to either the easy-medium or hard-medium group (n = 20 per group). Six participants who reported remembering more than 30% of the false details were replaced.

Materials, design, and procedure. The materials and procedure were the same as in Experiments 1 and 2. The easy-then-medium group was presented with a block of 15 easy details, followed by a block of 15 medium details. The hard-medium group was presented with a block of 15 hard details, followed by a block of 15 medium details. Five false details were also presented within each block. The order of details within each block was freshly randomized for each participant.

Results

Effect of test-list context on medium details. As Figure 4 shows, medium details were 14% more likely to be recognized when preceded by a block of hard details rather

than a block of easy details (.62 vs. .48), F (1, 19) = 8.26, MSE = .02, η^2 = .89. The presence of hard details in the first block also resulted in 18% more remember judgments for medium details (.38 vs. .20), F (1, 19) = 10.13, MSE = .03, η^2 = .91, but no difference in the rate of know judgments (.24 vs. .28), $F < 1$. This pattern is identical to that found in Experiment 2. The lack of test-list context effects on knowing in Experiments 2 and 3 contrasts with the inverse pattern between remember and know as a function of test-list context in Bodner and Lindsay (2003).

Effect of test-list context on false recognition. In contrast to Bodner and Lindsay (2003) and Experiment 2, the blocked list context affected false alarm rates as well. The hard-then-medium group had 15% higher rates of false recognition than the easy-then-medium group in the second block (.32 vs. .17), F (1, 19) = 6.41, MSE = .04, η^2 = .87 and were 13% more likely to assign know judgments to false details (.25 vs. .13), F (1, 19) = 4.51, MSE = .03, η^2 = .82. However, assignment of remember judgments to the medium block of false details was not influenced by whether they were preceded by hard or easy details (.07 or .04), F (1,19) = 1.10, MSE = .01, p = .43, η^2 = .52. Furthermore, the hard-then-medium group was as likely to assign know judgments to correct or false details (.24 vs. .25), $F < 1$, suggesting that they were unable to use know judgments effectively. The influence of the test-list context on discrimination and bias scores are considered below.

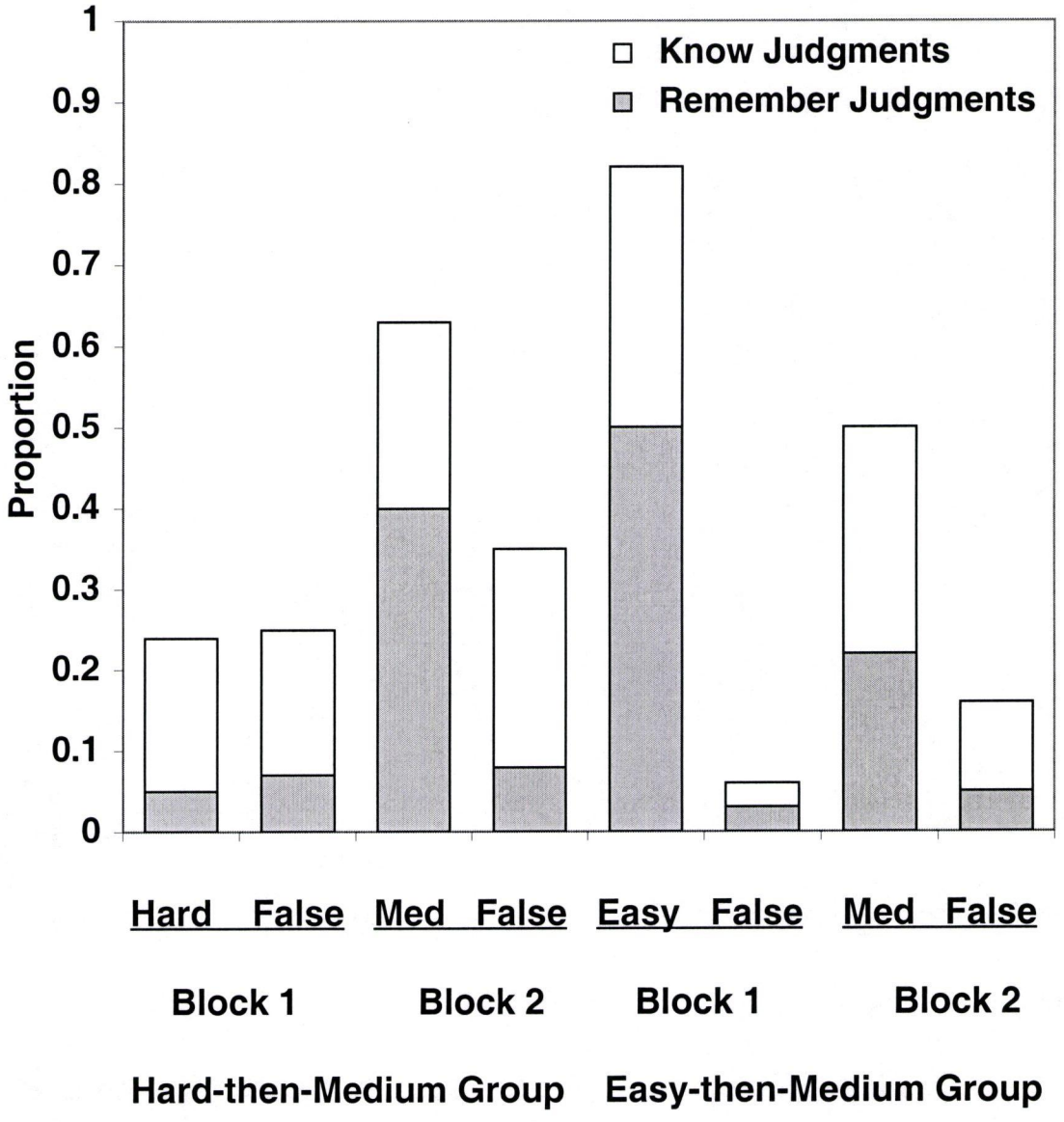

Figure 4. Mean proportion of remember and know judgments for studied and false details in the hard-then-medium and easy-then-medium groups in Experiment 3. The sum of remember and know judgments estimates overall recognition.

Effect of test-list context on discrimination. The discrimination scores for recognition, remember, and know judgments for the medium block of details are shown in Figure 5. The discrimination scores for recognition for the hard-then-medium and easy-then-medium groups were equal (.30 vs. .31), $F < 1$, suggesting that the differences

in correct recognition between the two groups was the result of the hard-then-medium group not being able to discriminate between correct and false details when using know judgments. However, the effect of test-list context on remember judgments remained, as discrimination scores for remember judgments were still 15% higher in the hard-then-medium group than in the easy-then-medium group (.31 vs. .16), F (1, 19) = 5.76, MSE = .04, η^2 = .85. There was also a marginal difference in discrimination scores for know judgments to the medium details for the hard-then-medium group versus the easy-then-medium group (-.01 vs. .15), F (1, 19) = 3.72, MSE = .07, p = .06, η^2 = .79. The equivalent corrected recognition for medium details between the hard-then-medium and easy-then-medium groups, coupled with an increase in remember scores/judgments, supports the functional account of recognition.

Effect of test-list context on bias. The bias scores revealed that the hard-then-medium group were less conservative than the easy-then-medium group about claiming to recognize medium details (.45 vs. .22), F (1, 19) = 13.36, MSE = .04, η^2 = .93, and were also less conservative when assigning know judgments (.22 vs. .13), F (1, 19) = 4.13, MSE = .02, η^2 = .81. However, both groups were equally conservative when assigning remember judgments to medium details (.08 vs. .05), F (1,19) = 1.14, MSE = .01, p = .29, η^2 = .53. Thus, the test-list context effects on remember judgments were again due to differences in discrimination, not bias, as in Experiment 2.

Discussion

The blocked design of Experiment 3 allowed for an examination of how the initial presentation of an easy or hard block of details influenced recognition judgments for a subsequent block of medium details. Higher recognition rates and remember rates for the

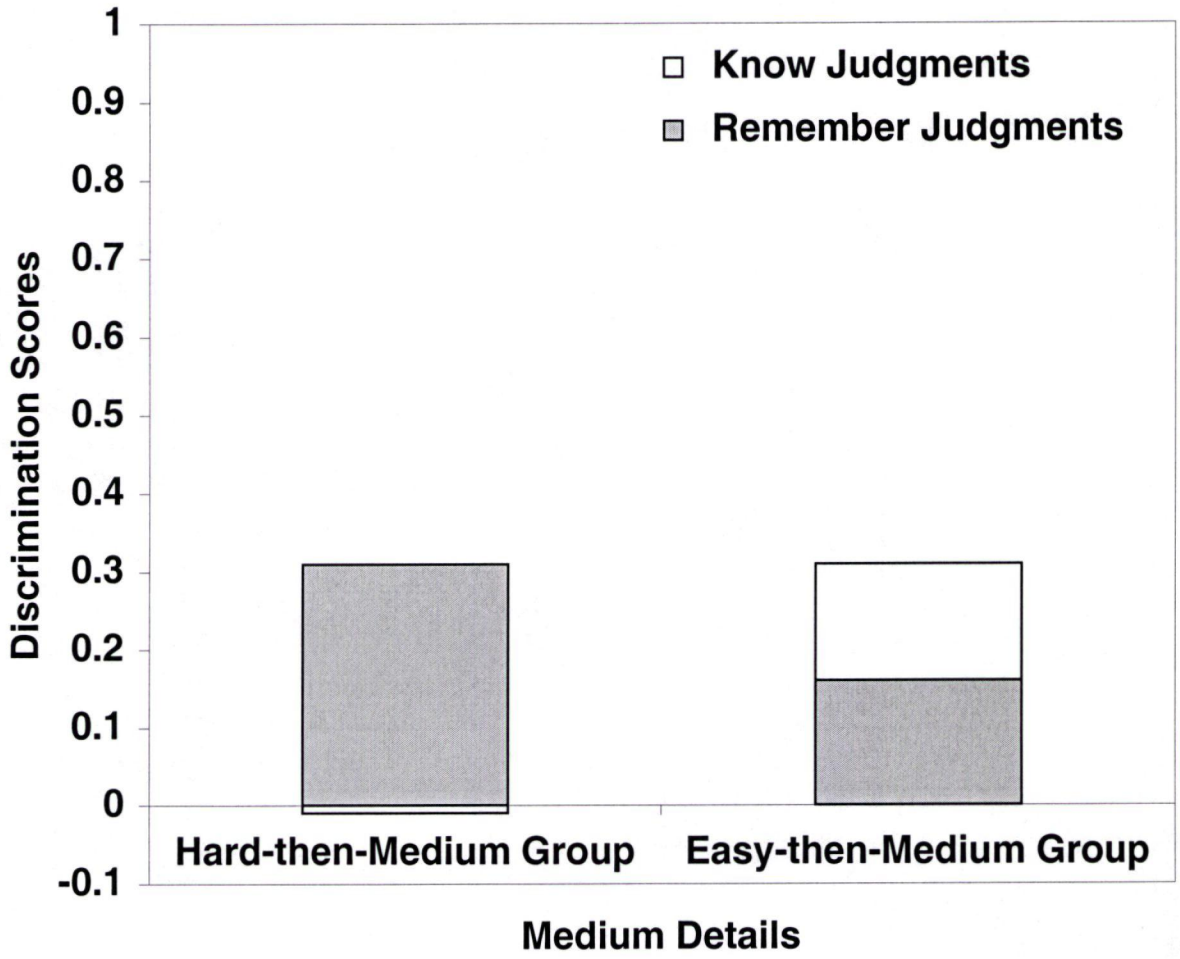

Figure 5. Mean discrimination scores for remember and know judgments for medium details in the hard-then-medium and easy-then-medium groups in Experiment 3. The sum of remember and know scores estimates overall recognition discrimination.

medium details were found in the hard-then-medium group, suggesting that the initial block of hard details caused participants to expect very little to no recollective experience for the medium block of details. Therefore, the increased recollective experience experienced for medium details may have been perceived as distinctive, resulting in more remember judgments (McCabe & Balota, 2005; Rajaram, 1998). In contrast, the easy-then-medium group may have decided that their recollective experiences for the medium

details did not contain enough information to warrant remember judgments, relative to the recollective experiences they had for the easy details in the first block.

Similar to Experiment 2, the difference in overall recognition between the two groups was attributable to different rates of remembering. Contrary to Experiment 2, false recognition and false know rates were higher in the hard-then-medium group, indicating a possible response bias between the two groups. However, the discrimination scores for overall recognition were equivalent once response bias was taken into account. This pattern of equivalent recognition scores, coupled with a higher rate of remembering in the hard-then-medium group, is consistent with the functional account of recognition. Bodner and Lindsay (2003) found equivalent recognition rates with a trade off in remember/know judgments, where medium LOP words were more likely to be assigned remember judgments when mixed within shallow LOP rather than deep LOP words.

Experiment 3 suggests that remembering is functionally defined at a global level. Participants appear to have defined remembering based on the details in the first block, and these functional definitions were carried forward to the second/(medium) block of details, resulting in different rates of remembering in the second block even though the details were the same for the two groups. If remembering experiences were defined on a local level (e.g., the experiences from the last few trials), then there should have been no differences in remember or know judgments for medium details between the two groups. Equivalent remember judgments for medium details the second block would have indicated that people changed the way they defined their remember experiences when their subjective recognition experiences changed. Although support was found for the global setting hypothesis, there were only 15 event details presented in each block, and

therefore it is possible that people would have redefined remembering if more event details had been included in the second block.

Experiment 4: Elimination of the Test-List Context Effect

Experiment 3 suggested that participants define remembering/knowing in a relatively global manner early on during the recognition test, rather than changing their functional definitions on a trial-by-trial or block-by-block basis. The purpose of Experiment 4 was to examine if providing information about the difficulty of the details presented in each block would cue participants to modify their remember/know definitions. The procedure for Experiment 4 was the same as Experiment 3, except that participants were informed of the difficulty of each block of details just before each block was presented. The functional account predicts that providing information about the difficulty of the details will encourage participants to redefine remembering, which should lead to similar definitions in the second block and thus similar assignment of remember judgments for the medium details in the second block.

Method

Participants. As in Experiment 3, forty participants were randomly assigned to either an easy-then-medium or hard-then-medium group (n = 20 per group). Nine participants who made remember judgments to 30% or more of the false details were replaced.

Materials, design, and procedure. Experiment 4 was identical to Experiment 3, except participants were informed about how difficult the details in each block would be to recognize. Before any details were presented, the easy-then-medium group was informed that the first block of details would be very easy recognize, and the hard-then-

medium group was informed that the first block of details would be very difficult to recognize. After the presentation of the first block of details, participants in the easy-then-medium group were told that the next block of details would be *more difficult* to recognize (to encourage the adoption of a more lenient definition of remembering relative to the first block), and the hard-then-medium group was informed that the next block of details would be easier to recognize (to encourage the adoption of a more conservative definition of remembering relative to the first block).

Results

Effect of test-list context on medium details. Figure 6 shows that the hard-then-medium and easy-then-medium groups had equivalent overall recognition rates (.59 vs. .54), F (1, 19) = 1.17, MSE = .03, p = .29, η^2 = .54, remember judgments (.29 vs. .31), F < 1, and know judgments (.31 vs. .23), F (1, 19) = 1.93, MSE = .17, p = .17, η^2 = .66, for the medium details. The test-list context effects found in Experiment 3 were eliminated when information was provided about the difficulty of the details to be experienced in each block.

Effect of test-list context on false recognition. The hard-then-medium and easy-then-medium groups showed equivalent rates of false recognition (.28 vs. .20), F (1, 19) = 2.85, MSE = .02, p = .10, η^2 = .74, false remember judgments (.10 vs. .06), F (1, 19) = 1.64, MSE = .01, p = .21, η^2 = .62, and false know judgments (.18 vs. .14), F (1, 19) = 1.64, MSE = .01, p = .21, η^2 = .62, for the medium details.

Effect of test-list context on discrimination. Discrimination scores (see Figure 7) were computed for comparability to Experiment 3. Medium details in the hard-then-medium and easy-then-medium groups led to equivalent recognition scores (.31 vs. .37),

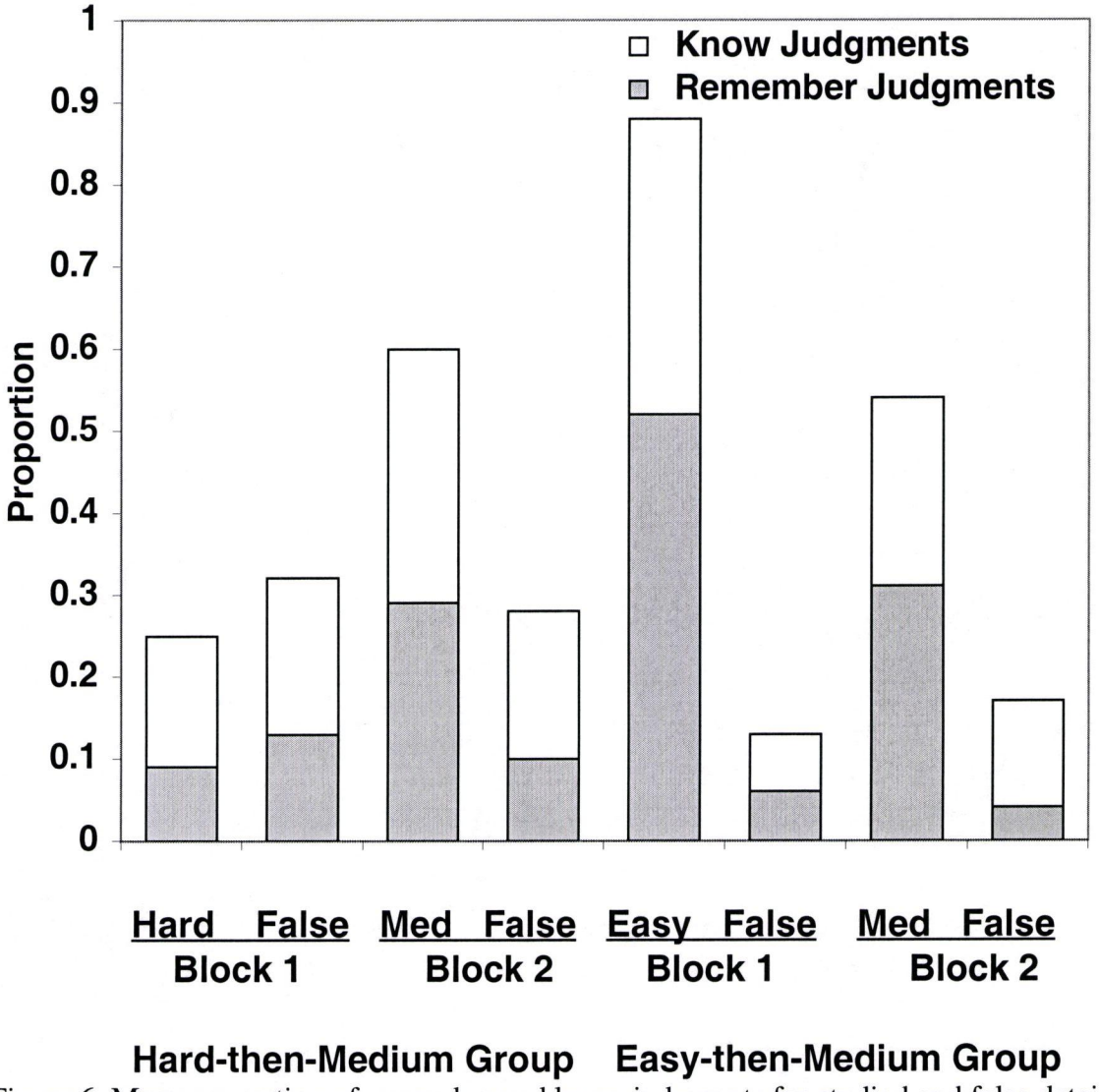

Figure 6. Mean proportion of remember and know judgments for studied and false details in the hard-then-medium and easy-then-medium groups in Experiment 4. The sum of remember and know judgments estimates overall recognition.

$F < 1$, know scores (.13 vs. .10), $F < 1$, and remember scores (.19 vs. .27), F (1, 19) = 1.71, $p = .20$, $MSE = .04$, $\eta^2 = .63$.

Effect of test-list context on bias. The bias scores revealed no differences between the hard-then-medium and easy-then-medium groups overall (.36 vs. .22), F (1, 19) = 1.61, $p = .22$, $MSE = .05$, $\eta^2 = .62$, or in terms of remember judgments (.10 vs. .06), F

(1,19) = 1.14, *MSE* = .01, $p = .29$, $\eta^2 = .53$, and know judgments (.18 vs. .11), F (1, 19) = 1.21, *MSE* = .04, $p = .28$, $\eta^2 = .55$.

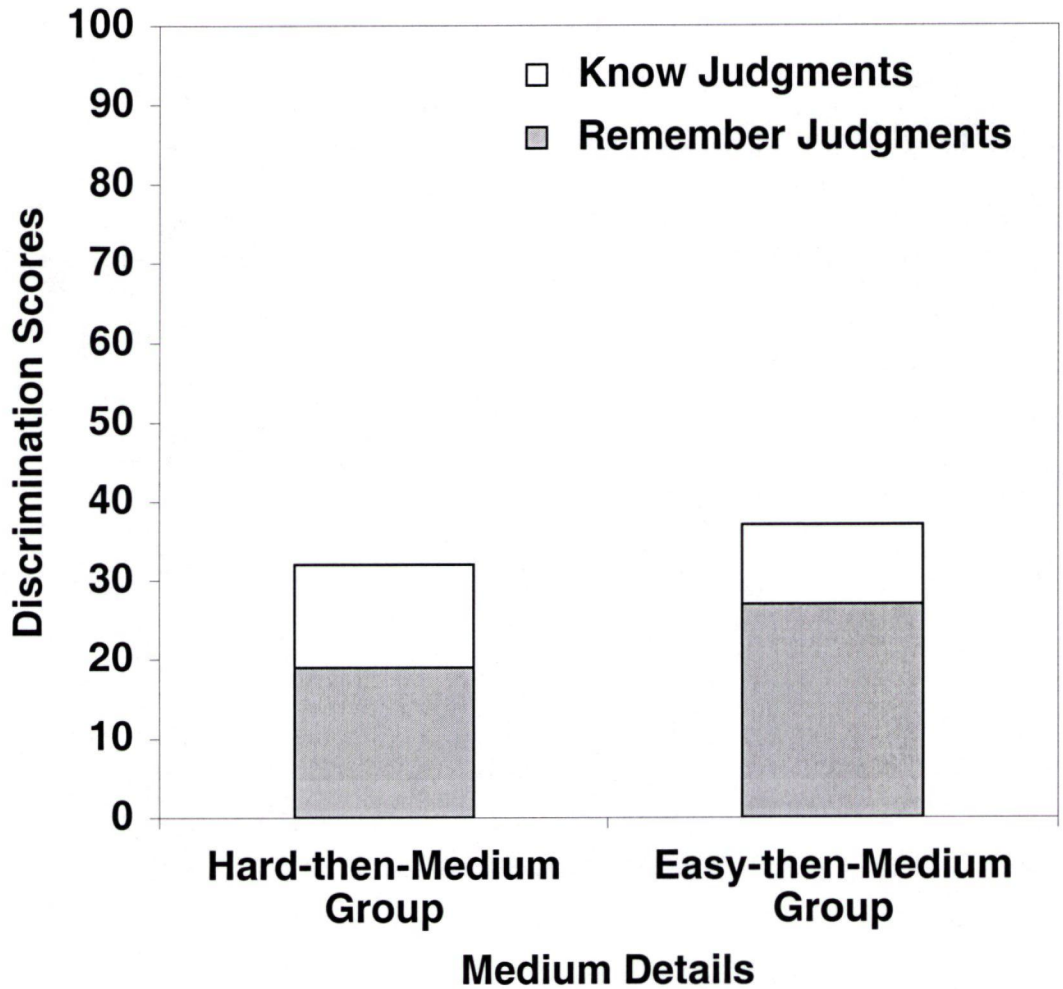

Figure 7. Mean discrimination scores for remember and know judgments for medium details in the hard-then-medium and easy-then-medium groups in Experiment 4. The sum of remember and know scores estimates overall recognition discrimination.

Discussion

The attempt at eliminating the blocked test-list context effects by providing information about the difficulty of the details in each block was successful. Providing information about the difficulty of the details in second block relative to the first block

caused people to redefine remembering in the second block such that the two groups' definitions coincided enough to produce equivalent rates of remember and know judgments.

Comparisons Between Experiments 3 and 4

To better understand why providing information about detail difficulty eliminated the context effects found in Experiment 4, a 2 (Experiment: 3 vs. 4) x 2 (Group: hard-then-medium vs. easy-then-medium) between-groups ANOVA was performed on discrimination scores for remember judgments, the measure in which the context effects were noted in Experiment 3 (and in Experiment 2). The interaction between experiment and group in the discrimination scores was significant, F (2, 38) = 6.96, MSE = .04, η^2 = .78. Pairwise comparisons revealed that providing instructions increased remember discrimination scores for the easy-then-medium group (.27 vs. .16), F (1, 38) = 5.00, MSE = .02, η^2 = .83, but decreased remember discrimination scores for the hard-then-medium group (.19 vs. .31), F (1, 38) = 2.89, MSE = .06, p = .09, η^2 = .74.

The decrease in discriminability for remember judgments for the hard-then-medium group in Experiment 4 was likely the result of their better ability to discriminate between studied and false details when making know judgments. Without instructions, participants in the hard-then-medium condition assigned know judgments to studied and false details at similar rates in Experiment 3 (.24 vs. .25), $F < 1$, but with instructions, participants in this condition assigned more know judgments to studied details in Experiment 4 (.31 vs. .18), F (1, 19) = 8.35, MSE = .02, η^2 = .89. Thus, when participants in the hard-then-medium group were informed that the second block of medium details would be easier to recognize, it appears that they set higher standards for classifying their

recognition experiences as remembering and knowing, resulting in better discrimination when using know judgments.

General Discussion

The present set of experiments extended Bodner and Lindsay's (2003) findings with word lists by showing an influence of the test-list context on memory for the details of an event. The same set of medium details were more likely to be classified as "remembered" if they were mixed with a set of hard details (Experiment 2) or presented after a block of hard details (Experiment 3), relative to conditions where the medium details were mixed with or presented after a block of easy details. Although rates of remembering were affected by the test-list context, overall recognition rates for medium details were similar across test contexts when corrected scores were used to subtract out differences in guessing rates (i.e., false alarms). Thus, consistent with a functional account of recognition (Bodner & Lindsay, 2003; Grupposo et al., 1997), context affected how participants interpreted their recognition experiences, but it did not affect their likelihood of having recognition experiences. In essence, the difficult test-list contexts (hard/medium, hard-then-medium) caused participants to adopt a more liberal definition of remembering than the easy test-list contexts (easy/medium, easy-then-medium). The same medium details studied in the same manner were more likely to "stand out" and hence to foster remember judgments in the "hard" context than in the "easy" context.

According to the distinctiveness-fluency account, remember experiences are driven by perceptions of distinctiveness, and know judgments are driven by perceptions of fluency (Rajaram, 1993, 1996, 1998; Rajaram & Geraci, 2000). To accommodate the present result, this account would suggest that the hard test-list contexts led to greater

perceived distinctiveness and hence higher rates of remember judgments for medium details relative to the easy test-list contexts (e.g., easy/medium group, easy-then-medium group). McCabe and Balota's (2005) "relative distinctiveness hypothesis" extended this distinctiveness-fluency account by focusing on the *perceived* distinctiveness of items relative to other items on the test. This account suggests that the assignment of a remember or know judgment for an item depends on people's expectation that their recognition experience for that item will be distinctive, relative to that produced by other items on the test list. On this view, the test-list context affects people's expectancies about their recognition experiences, and can cause the same experience to be perceived as more distinctive in one context relative to another.

Applied to the present study, McCabe and Balota's (2005) analysis suggests that participants given the hard contexts did not expect to have very distinctive recognition experiences, and hence their recognition experiences for the medium items often exceeded their expectation, thus increasing their reports of remember judgments. In contrast, participants given the easy contexts expected quite distinctive recognition experiences, and hence their recognition experiences for the medium items would often fail to meet this expectation, thus reducing their reports of remember judgments. The relative distinctiveness hypothesis can also account for the elimination of the test-list context effect when information about detail difficulty in each block was provided. This information would change participants' expectations regarding the distinctiveness of their recognition experiences in the second block. For example, informing the hard-then-medium group that the second block of details would be easier to recognize caused them

to not be surprised by the increased distinctiveness of their recognition experiences (and vice versa for the easy-then-medium group).

The test-list context effects in Experiments 2 and 3 pose a significant challenge to the signal detection account. Given that encoding processes were identical for both groups, the test-list context manipulations should have affected response bias rather than discrimination. In contrast, the functional account can explain why the test-list context manipulations affected discriminability rather than response bias. One possibility is that the hard-context groups became more sensitive to diagnostic aspects of their recollections for the medium details, resulting in them constructing more effective functional definitions of remembering than the easy-context groups. Similarly, according to the relative distinctiveness hypothesis, medium details in the hard context would be expected to contain very little recollected information. This low expectancy would be exceeded by the distinctiveness of the actual recognition that accompanied the medium details, resulting in improved discrimination when making remember judgments.

To explain the test-list context effects in Experiments 2 and 3, proponents of the memory systems account would have to claim that the same medium details were retrieved from the episodic memory system in the hard context, but were retrieved from a semantic/perceptual memory system in the easy context. Although this is possible, it would seem rather odd for the easy-context groups to not rely on the episodic memory system more than the hard-context groups, given that the former group would tend to have richer recognition experiences during the test.

Finally, the dual-process account fails to explain why the different test-list contexts would produce different rates of remember judgments. According to this

account, remembering is a pure measure of recollection, which is a threshold process that should not be affected by the test-list context.

Even though the functional account of recognition appears to describe the current findings well, it is not clear whether the different test-list context groups used different bases when defining remember and know, as Bodner and Lindsay (2003) found. Bodner and Lindsay were able to show that qualitatively different bases were emphasized at test when groups had performed different encoding tasks at study. In the present study, however, the encoding event was the same for both test-list context groups. Therefore participants may not have emphasized different encoding attributes when making remember/know judgments. However, it remains possible that different attributes were nonetheless used in the different groups, and future research could ask participants to report the bases on which they make their remember/know judgments to determine whether this is the case. Alternatively, participants might simply experience quantitatively "more" recollection in the hard contexts than in the easy contexts, more in line with a signal detection perspective.

In Experiment 3 (blocked design), the increased rates of remember judgments assigned to medium details after a block of details relative to a block of easy details suggests that remembering is functionally defined at a global level. However, the elimination of test-list context effects in Experiment 4 revealed that this is not always the case. That is, providing information about the difficulty of the details just prior to a block appears to have caused participants to redefine remembering in the second block such that the two groups came to set similar functional definitions. Thus, functional definitions can be modified if people are provided with information that encourages them to do so.

Conclusions

The present experiments reveal that test-list context effects previously found with word lists can be extrapolated to memory for an event. The corrected scores obtained in both mixed and blocked test-list contexts suggest that the amount of recollected information is not much affected by the test-list context. Instead, the test-list context influences people's interpretation of whether they feel they are having an episodic recollection for an event detail or not. This finding has potentially important implications for legal and forensic interview situations. For example, based on the results of Experiment 3, if eyewitnesses are asked a series of questions about event details that they do not remember (hard details), then they may report remembering a second set of event details that are not as difficult to remember (medium details). But if the same eyewitnesses were first presented a series of questions about very memorable event details (easy details), then they may be less likely to report remembering the second set of details.

For example, eyewitnesses to a hit-and-run accident may claim to remember that a suspect had blonde hair if first asked a difficult question about a license plate number, whereas they may claim that blonde hair "seems familiar" if first asked an easy question about the colour of the car. Differences in witnesses' judgments about their memory for various details of the event could influence the way a legal investigation is carried out. Future research should examine whether test-list context effects also occur in cued-recall tasks, which are more likely to be used in actual investigations (e.g., "What colour was the suspect's hair? Do you remember or know this detail?" vs. "The suspect's hair was blond. Do you remember or know this detail?").

The outcomes of legal and forensic investigations have a considerable impact on peoples' lives. Therefore, it is important to be aware of all of the factors that influence the way people report their memory for an event. The test-list context effects observed in the lab may be exacerbated in a courtroom or crime scene situation where witnesses may be nervous, intimidated, and/or anxious to please an authority figure. Such feelings may encourage witnesses to increase the weight they assign to their recollective experiences. People may also overcompensate for their lack of knowledge about difficult event details by embellishing their judgments about other, less obscure details from an event. The embellishment of one's memory for certain event details could potentially lead to the imprisonment or perhaps even the execution of an innocent person.

References

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48,* 563-580.

Brown, J. M. (2003). Eyewitness memory for arousing events: Putting things into context. *Applied Cognitive Psychology, 17,* 93-106.

Fisher, R. P., & Geiselman, R. E., & Amador, A. F. (1989). Memory-enhancing techniques for investigative interviewing: The cognitive interview. Charles C. Thomas: Springfield, IL.

Gardiner, J. M., & Java, R. I. (1993). Recognition memory and awareness: An experiential approach. *European Journal of Cognitive Psychology, 5,* 337-346.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2,* 259-278.

Hirshman, E. (1995). Decision processing in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 302-313.

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology, 18,* 711-726.

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1534-1555.

Kahn, H. (Producer/Director). (2004). Water's Edge [Motion picture] Ch. 5. United States: Wicked Seasons Productions, Inc.

Kebbell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology, 87,* 653-662.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585-589.

Loftus, E. F. (1992). When a lie becomes memory's truth: Memory distortion after exposure to misinformation. *Current Directions in Psychological Science, 1,* 121-123.

Marquis, K. H., Marshall, J., & Oskamp, S. (1972). Testimony validity as a function of question form, atmosphere, and item difficulty. *Journal of Applied Social Psychology, 2,* 167-186.

Marsten, W. M. (1924). Studies in testimony. *Journal of the American Institute of Criminal Law, 15,* 5-31.

McCabe, D. P., & Balota, D. A. (2005). Context effects on remember-know judgments: The influence of relative word frequency at study and test. (In Review).

McCloskey, M. & Zaragoza, M. (1985). Misleading post-event information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114,* 1-16.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experiemental Psychology: Learning, Memory, and Cognition, 17,* 855-874.

Rajaram, S. (1993). Remembering and knowing: Two ways of assessing the personal past. *Memory & Cognition, 21,* 89-102.

Rajaram, S. (1996). Perceptual effects on remembering: Two means of access to the personal past. *Memory & Cognition, 21,* 89-102.

Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review, 5,* 71-78.

Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1070-1074.

Ratcliff, R., Clark, S., & Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163-178.

Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99, 518-535.*

Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition, 22,* 63-69.

Smith, S. M. (1988). Environmental context-dependent memory. In Thompson D. M., Davies, G. M. (Eds.), *Memory in Context: Context in Memory* (pp. 13-34). John Wiley: New York.

Snodgrass, J. G. & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.

Stern, L. W. (1910). Abstracts and lectures on the psychology of testimony and on the study of individuality. *American Journal of Psychology, 21,* 270-282.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1-12.

Whittlesea, B. W. A. (2002). Two routes to remembering and another to remembering not. *Journal of Experimental Psychology: General, 131,* 96-115.

Winkielman, P., & Schwartz, N. (2001). How pleasant was your childhood? Beliefs about memory shape inferences from experienced difficulty of recall. *Psychological Science, 12,* 176-179.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441-517.

Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 345-355.