

ABSTRACT

Eyewitness Testimony in Civil Litigation: Retention, Suggestion, and Misinformation in Product Identification

Jonathan Trent Terrell, Ph.D.

Mentor: Charles A. Weaver, III, Ph.D.

Expert testimony in eyewitness memory cases is now common in criminal cases. However, eyewitness testimony is also critical in civil litigation, particularly in product liability cases involving alleged exposure to toxic substances like asbestos. Witnesses in these cases must recall specific brands of products that may have been used decades earlier. The present experiments investigate eyewitness memory for product brand names seen in videos of cooking shows and news reports. Although memory was reasonably accurate at brief delays, within a week recognition rates for the brand names dropped to scarcely above chance; nearly half of these delayed selections were of the most familiar (but unseen) brands. Subtle and inaccurate post-event suggestions embedded in questionnaires produced robust false alarm rates—nearly 70% of responses when the most popular brands were suggested. Refreshing with photographs of products also had a significant impact on identifications—when two brands were shown during refreshing, participants identified one of these between 75 and 90% of the time, regardless of the accuracy of the suggestions. Further, a digital editing program was used to produce

photographs of products that do not exist—even these implausible products were identified over one quarter of the time following refreshing. Finally, refreshing designed to be neutral had little effect on identification, suggesting that the primary mechanism of photo refreshing was suggestion rather than true memory jogging. Metamemory measures revealed that confidence in false alarms often matched or exceeded that attributed to correct identifications, a finding that is particularly disconcerting when the effect of witness confidence on jurors' perception of reliability is considered. The implications of these findings for civil law are discussed.

Eyewitness Testimony in
Civil Litigation: Retention, Suggestion, and
Misinformation in Product Identification

by

Jonathan Trent Terrell, B.A., M.A.

A Dissertation

Approved by the Department of Psychology and Neuroscience

Jaime L. Diaz-Granados, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by the Dissertation Committee

Charles A. Weaver III, Ph.D., Chairperson

Jaime L. Diaz-Granados, Ph.D.

Wade C. Rowatt, Ph.D.

Matthew S. Stanford, Ph.D.

James A. Marcum, Ph.D.

Accepted by the Graduate School
May 2008

J. Larry Lyon, Ph.D., Dean

Copyright © 2008 by Jonathan Trent Terrell

All rights reserved

TABLE OF CONTENTS

List of Figures	vii
List of Tables	viii
Acknowledgments	ix
CHAPTER ONE	1
Introduction	1
The Impetus for Research: Product Liability Cases	1
Misinformation and Metamemory	6
The Present Research	9
CHAPTER TWO	11
Participants, Materials, Methodology, Results, Discussion	11
Experiment One	11
Participants	11
Materials	12
Method	13
Results	14
Analyses	14
Accuracy Measures	14
Confidence/Accuracy Relationship	16
Discussion	18
Experiment Two	20

Experiment 2a	21
Participants	21
Materials	21
Method	22
Results	23
Analyses	23
Accuracy Measures	24
Confidence	25
Confidence/Accuracy Relationship	25
Experiment 2b	26
Participants	26
Materials	26
Method	26
Results	26
Analyses	26
Accuracy Measures	26
Confidence	27
Confidence/Accuracy Relationship	27
Discussion	28
Experiment Three	32
Experiment 3a	33
Participants	33
Materials	33

Method	33
Results	34
Analyses	34
Accuracy Measures	34
Confidence	35
Confidence/Accuracy Relationship	35
Experiment 3b	36
Participants	36
Materials	36
Method	36
Results	36
Analyses	36
Accuracy Measures	36
Confidence	37
Confidence/Accuracy Relationship	38
Discussion	39
Experiment Four	43
Participants	43
Materials	43
Method	44
Results	44
Analyses	44
Accuracy Measures	45

Confidence	47
Confidence/Accuracy Relationship	47
Discussion	48
Experiment Five	50
Participants	51
Materials and Method	51
Results	52
Analyses	52
Accuracy Measures	53
Confidence	53
Confidence/Accuracy Relationship	54
Discussion	55
CHAPTER THREE	58
General Discussion	58
Implications for Civil Law	68
Limitations	68
Final Conclusions	72
APPENDICES	74
Appendix A	75
Appendix B	76
Appendix C	77
Appendix D	78
References	79

LIST OF FIGURES

Figure 1	Confidence Calibration Curves for Hits, Experiment One	17
Figure 2	Confidence Calibration Curves in the Ten-Minute Delay, Familiar Misinformation Condition, Experiment 2	24
Figure 3	Confidence Calibration Curves in the One-Week Delay, Familiar Misinformation Condition, Experiment 2	29
Figure 4	Confidence Calibration Curves Following Refreshing with the Correct Product, Experiment 3b	38
Figure 5	Confidence Calibration Curves Following Refreshing with Exclusively Incorrect Products, Experiment 3b	39
Figure 6	Confidence Calibration Curves, Different Brand/Different Container Condition, Experiment 4	48
Figure 7	Confidence Calibration Curves for Hits, Experiment 5	55

LIST OF TABLES

Table 1	Means, Experiment One	16
Table 2	Means, Experiment 2a	23
Table 3	Means, Experiment 2b	28
Table 4	Means, Experiment 3a	35
Table 5	Means, Experiment 3b	37
Table 6	Means, Experiment 4	46
Table 7	Means, Experiment 5	54

ACKNOWLEDGMENTS

Thanks first of all to those who have served on my committee throughout this process—Jim Diaz-Granados, Wade Rowatt, Matthew Stanford, Greg Garrett, and James Marcum. Your insights and refinements have been helpful, and your willingness to act as my committee was much appreciated. Thanks primarily to Chuck Weaver, who took me under his wing and was the ideal mentor from day one. Thanks to more undergraduates than I could list within the confines of graduate school formatting. Thanks to Amanda Holmes and the rest of my co-graduate students, and thanks to Nancy Ulman and Rita Massey, who aided in all things technical with a glad heart. Thanks most of all to my family and my wife Megan, whose unwavering faith (some of it blind) and love helped me make it to this point.

CHAPTER ONE

Introduction

An increasing awareness of memory's reconstructive nature has prompted a thorough reevaluation of the usefulness of eyewitness memory testimony in the criminal courtroom (Wells & Loftus, 2003). The Department of Justice has created a thorough series of guidelines for refreshing recollection in criminal cases (see U.S. DOJ, 1999), and memory experts are commonly called upon to make clear to jurors in such cases that "memory is not a video tape." April 23, 2007 marked the 200th criminal conviction exonerated by DNA evidence in the United States. According to www.innocenceproject.org, over 75% of these wrongful convictions involved faulty eyewitness testimony. Collectively, these 200 people spent a total of 2,475 years in prison. Only four years ago, Loftus (2004) reviewed this growing problem and how it happens on the heels of the 100th DNA exoneration. Clearly, scrutiny of eyewitness accuracy is becoming prominent in the courtroom, where testimony based on the recollections of witnesses often plays a vital role in determining innocence or guilt. Often the memory of a single individual represents the crux of the prosecution's case, and thus considerable research has been conducted to help evaluate the tenability of witnesses' reports.

The Impetus for Research: Product Liability Cases

Eyewitness memory research has focused almost exclusively on central details obviously relevant to future questioning, such as faces, car colors, and tools appearing in

a robbery (Loftus et al., 1989). Comparatively little effort has been expended on studying memory for details insignificant at the time of encoding, such as product brand names. However, long-term memory for such things is growing increasingly pertinent in the courtroom, as the number of lawsuits hinging on claims from the remote past (25-40 years) grows steadily (see Biederman et al., 1998). In particular, cases involving exposure to asbestos brought by those suffering from asbestiosis or mesothelioma often rely heavily on a victim's recollection of exposure to the asbestos-containing product manufactured or distributed by the companies being sued. Virtually no research has been conducted on long-term memories for details inconsequential at the time of acquisition in such cases, such as recollections of the brand names of joint compound or modeling clay mixes containing asbestos (see Colby & Weaver, 2006; Krug & Weaver, 2005; and Terrell & Weaver, in press, for exceptions). In most cases, written records that might confirm the presence or absence of asbestos-containing materials are unavailable, and the witness's recollections of working in its presence become the essence of the case.

Currently, civil law places few guidelines or restrictions on the manner of refreshing allowed by attorneys in such product identification situations. The problem posed by an absence of such restrictions is best explained with the following example (see Robinson vs. Chesterton, 2001). The Narcolite Cement Company produced two products: Narcolite Castable (thought to be asbestos-free), and Narcolite Gunning Mix (thought to contain trace amounts of asbestos). In 1996, the first twenty witnesses deposed in Cuyahoga County, Ohio recalled exposure to Narcolite Gunning Mix. In 2000, new information revealed that it was actually Castable that contained asbestos, while Gunning Mix was asbestos-free. All but one of the most recent twenty deponents

have identified Castable, while some initially remembering Gunning Mix have switched over to Castable in later depositions.

Brickman (2004) offers detailed descriptions of some of the refreshing tactics that can produce patterns of identification like that described above. Witnesses are often suggestively shown photographs of products, and encouraged to make certain identifications with remarks like, “Others from that site remembered seeing this product, what about you?” Further, products manufactured by companies no longer in business are often excluded from refreshing arrays, all but guaranteeing that any identifications made will provide a tenable basis for legal action.

Consider another example. During a recent asbestos trial (Hampel, 2005), the fifty-five-year-old daughter of a mesothelioma victim claimed that she remembered the brand of joint compound her father had been using during a distant project, and was able to describe its container in some detail. When a defense attorney asked her if she recalled the type of paint her father had been using while working on the same project, she replied with an incredulous “No,” and then added, “It was thirty-five years ago.” Interactions with the prosecuting attorneys had perhaps led this individual to believe in an unlikely recollection, one she herself refuted having for other similar products. That the capability to recall a brand of paint from thirty-five years ago should not differ from the capability to recall a brand of joint compound used thirty-five years ago seems to have escaped this witness. More importantly, notions such as these are often unappreciated by juries, as well.

The impact of product liability litigation is enormous. A recent report issued by the Rand Corporation (Carroll et al., 2005) reported that asbestos litigation has cost more

than \$70 billion in the United States to date, with final estimates as high as \$250 billion. Of this total, according to the Rand report, less than half is eventually paid to claimants.

The existing research on long-term memory suggests that recollections of products used so remotely may be unreliable, and that suggestive refreshing by attorneys (see Biederman et al., 1998; Brickman, 2004) may therefore act to create new memories rather than facilitate retrieval of old ones. The Deese-Roediger-McDermott false memory paradigm reveals the ease with which false memories can be created (see Roediger & McDermott, 1995). Here, participants study a list of words related to one critical, but non-presented, word, and later tend to incorrectly recall this critical lure as having been on the list. Brainerd and Reyna (2004) posit that these errors result from examining the overall “gist” of the event being remembered, a technique that draws from semantic associations as much as episodic recall. This helps to explain the prominence of familiarity bias in product identification, a phenomenon discussed further below.

Two recent articles illustrate the powerful effects exerted by retrieval cues (Lindsay, Hagen, Read, Wade, & Gary, 2004; Wade, Garry, Read, & Lindsay, 2002). Wade et al. (2002) created a false photograph of a childhood event. They took an existing photograph of a parent and child riding in a hot air balloon, and digitally inserted the faces of the experimental participants (as children) and their parents over the existing faces. Half of the individuals shown this false photograph created a complete or partial memory for this fictitious event.

In a second study, Lindsay et al. (2004) showed participants true photographs as a way of soliciting childhood memories. In addition to questioning participants about a real event, however, Lindsay et al. asked about a false event (placing “slime” in their

teacher's desk). When participants were shown class photographs that coincided with this fictitious event, the rate of false memory reports were twice as high as they were when participants were not shown photographs, even though these were real photographs. Even real photographs can be reconstructed into false memories. Experiment 4 will examine the effect of digitally-manipulated photographs on subsequent product identifications.

Garry and Wade (2005) recently demonstrated that narratives about a plausible (but imaginary) event—a hot air balloon ride just as described previously—were even more effective than photographs at inducing illusory memories. The researchers generated a very brief (45 words) description of a hot air balloon ride and presented it with 3 other (true) childhood event descriptions. An astonishing 82% of their participants “recalled” at least some details from this fictitious event. Garry and Gerrie (2005) provide an excellent overview of this research.

Origins of Long-Term Memory Research

The transience of long-term memories has long been known. Ebbinghaus (1913/1885) first demonstrated the general inability of humans to remember over long periods of time with his pioneering study that established the now-famous forgetting curve. Ebbinghaus used himself as the subject for his study, spending months memorizing series of non-sense syllables to see how well he could later remember them. He found that his memory for the syllables quickly declined, dropping below 50% in less than an hour. The curve indicates most of his forgetting occurred within 24 hours, and only slight decreases occurred over subsequent days and months. Similar patterns of forgetting have been observed for numerous classes of information,

including names and faces of high-school classmates (Bahrick, Bahrick & Wittlinger, 1975) and Spanish vocabulary words learned fifty years previously (Bahrick, 1984). The negligible amount of additional forgetting occurring after one week therefore appears to justify the use of such a delay for conducting research on the limits of long-term memory, and as such a delay of this length is used in many of the experiments described below.

Misinformation and Metamemory

One of the most frequently replicated and commonly used paradigms to study erroneous memory uses misinformation, whereby false information introduced following the event of interest alters the witness' memory of the event (Loftus et al., 1989). For example, an eyewitness may remember seeing a yield sign initially, but if an investigator asks a question about a stop sign, the witness may integrate this assertion into their memories (Loftus et al., 1978).

While Loftus posits that conflicting post-event information eradicates original memory traces, others contend that the faulty suggestions only displace them at the time of retrieval, producing inaccurate recollections. These are called source attribution errors (Chandler et al., 2001; Frost & Weaver, 1997; Lindsay, 1990). Both theories have been successful in describing the effects of inaccurate post-event suggestions on subsequent recollections. Schacter (1996) describes an interesting instance of mistaken eyewitness memory in which a woman who had been raped later identified her assailant from a standard police lineup. Her chosen attacker, however, had an iron-clad alibi: He had been giving a live television interview at the time of the attack. In fact, the woman had been watching the interview before the intrusion, and later confused the man she had

been watching on television for the man who brutally attacked her. This bizarre instance represents a rather extreme example of a source attribution error, whereby witnesses correctly remember exposure to something, but misattribute its source. Some evidence suggests that these errors can occur even with the original memory trace intact (see McCloskey & Zaragoza, 1985). Here, participants defer to post-event suggestions over their memory in an attempt to maximize the likelihood of correct responding—a tactic that emerges when the memory task is particularly difficult. This deliberate strategy likely factors into the phenomena described below.

Much recent research has been devoted to metamemory—the ability of people to make estimations about the content, functioning, and capabilities of their memory system (Costermans, Lories & Ansay, 1992). The present research focuses on participants’ estimations of accuracy immediately after answering memory questions. These judgments are important in both civil and criminal law, as the confidence with which witnesses report their recollections is often perceived by jury members to be positively correlated with the accuracy of their memories (Winningham & Weaver, 2000). However, the relationship between confidence and accuracy is modest, at best (Kelley & Lindsay, 1993; Koriat, 1995; Loftus et al., 1989). While there is a modest correlation between confidence and accuracy of answers provided by control participants not exposed to misinformation, this correlation vanishes in participants who have been exposed to conflicting information. Their confidence remains high, despite the inaccuracy of their memories (Loftus et al., 1989).

The ease with which memories are recounted, or the “fluency” with which answers to questions are retrieved, may also affect the confidence people place in their

memories. Kelley and Lindsay (1993) exposed participants to incorrect information before asking them to answer general knowledge questions (for example, having them read a list of words including “leopard” before asking them which is the fastest animal on Earth). Participants quickly retrieved this false information and believed it to be true, answering with “leopard” instead of “cheetah.” Even if participants knew the correct answer, they may be swayed by the swiftness and fluency with which the indirectly suggested alternative pops into their minds. Similarly, Dunning and Stern (1994) argue that familiarity bias may result from the most familiar answer choices “popping out” at participants when completing a multiple choice memory test. In the experiments described below, participants will be asked to rate their confidence in the accuracy of each answer they provide using a scale ranging from 0% to 100%. This will allow comparisons to be drawn between the accuracy of memories and the confidence participants place in them.

Considerable research has been conducted on this ability of participants to predict their future performance, an exercise described in the literature as judgments of learning (JOLs) (Koriat, 1997). Many theories have been proposed to explain the basis for these judgments, one of the most prominent being that answers produced more quickly are assigned a higher JOL. There is a reliable inverse relationship between the rate of response generation and the corresponding JOL (Benjamin, Bjork & Schwartz, 1998; Matvey, Dunlosky & Guttentag, 2001), whether the judgment is made by the possessor of the memory or by an otherwise uninvolved person observing the decision making process. This suggests that at least some portion of the confidence assignment process is achieved indirectly via an analytical process not involving the strength of the memory

trace. The presumption, by the retriever or the observer, is that information that does not spring to mind immediately is likely not as well remembered as that which is retrieved with great fluency.

Overall, the literature supports the conclusion that confidence in memories should not be regarded as a reliable predictor of their accuracy, especially following the introduction of conflicting post-event information. The results reported below also support this conclusion. Loftus et al. (1989) assert that erroneous, misinformation-induced memories are subjectively as real to participants as memories based on actual events. This in itself should cast considerable doubt on the accuracy of eyewitness testimony following refreshing or persistent questioning by authorities.

The Present Research

The evidence documenting the considerable effects of inaccurate post-event information on memories is plentiful and undisputed. Having established the existence of such effects, much of the subsequent research in this area focused on the resolution of various theoretical explanations for the phenomenon. In this interest, the basic format of the research has hinged on standard paradigms that manipulate significant information about certain primary objects in a series of events, such as the color of a car or the kind of tool taken in a robbery. While much of the classic research on misinformation has also included questions regarding less consequential details, such as the brand of soda can present in a series of slides (McCloskey & Zaragoza, 1985), these studies have done little besides confirm the authors' suspicion that such memories are suspect and likewise susceptible to the effects of conflicting post-event information. Despite these

observations, there exists a relative paucity of research considering the factors that affect long-term memory for details that are trivial at the time of encoding.

The following experiments explored the effects of misinformation and other forms of suggestion—chiefly photo refreshing—on product identification. All of them utilized short films to present products to participants in a controlled and regulated manner. Experiment 1 examined the role of awareness at the time of encoding. One group of participants knew that a product identification test was forthcoming, while the other received no such warning. Experiment 2 explored the effects of written misinformation embedded in questionnaires. Participants were exposed to familiar products, unfamiliar products, or no products at all in questionnaires unrelated to product identification. In Experiment 3, participants were refreshed with pairs of three different combinations of product photographs—correct/familiar, correct/unfamiliar, and familiar/unfamiliar—before completing the identification test. Experiments 1 - 3 were each completed utilizing both a five-to-ten minute delay and a one-week delay between encoding and testing. Experiment 4 was similar to Experiment 3, but included photos of non-existent products created with PhotoShop. In Experiment 5, participants were refreshed neutrally with photos of all five products appearing on the identification questionnaire. Experiments 3 – 4 only afforded participants the opportunity to view photos of two eventual product choices.

CHAPTER TWO

Participants, Materials, Methodology, Results, Discussion

Experiment One

The first experiment sought to establish that the film provided adequate exposure to the products for them to be remembered subsequently while concurrently investigating the effects of various delays between witnessing the event and questioning. Half of the participants were informed of the purpose of the study, and as such knew that they would be tested on the brand names of the products. The other half were given no indication of the experiment's purpose, and were only asked to pay careful attention to the film and be prepared to answer some questions regarding it. Some participants had only a five minute delay before attempting to remember the brand names, while the remainder returned following a one week delay to attempt the same task. The variables in this initial study were therefore the length of the delay between film viewing and questionnaire completion, and the intentions of the participants during encoding.

Participants

Seventy-six psychology students (48 female, 28 male) aged 18-22 from Baylor University received extra credit for their participation. This experiment suffered a particularly high attrition rate, as a forced evacuation of the building in which the experiments were conducted occurred just before two of the follow-up sessions were scheduled to commence. As a result the data from several participants were lost, and some who did complete the follow-up session did so between eight and ten days after the

original event, rather than the intended seven. These participants were not removed from analysis.

Materials

Participants viewed various baking products by watching a short film resembling a cooking program. During the program, the chef prepared a chocolate chip cookie mix using sugar, flour, salt, baking powder, baking soda, eggs, and chocolate chips. Each product appeared twice during the film. The first exposure was a brief close-up of each product as the chef read through the necessary ingredients for the recipe. Each product filled the entire screen for around two seconds, and the products were shown in succession without other shots separating them. Later, the products were seen again as the chef used each of them to prepare the cookie mix. Though the products were not displayed as prominently in these shots, they were clearly visible and remained the central focus of attention. These second exposures lasted around eight seconds. The brand names were never spoken during the film, nor were they presented in any way other than on the containers themselves. The film's duration was just under five minutes.

An article about the use of film as a therapeutic tool (Suarez, 2003) was employed as a distractor task between watching the film and answering subsequent questionnaires. This topic was chosen to lend credence to the illusion that the study's primary focus was on the qualities of the film rather than the information it presented.

Immediately after reading the article, participants in the one week delay condition were given a brief questionnaire that pertained to the film's style (see Appendix A). This exercise was conducted primarily to disguise the fact that the purpose of the study was to test their long-term memory for the same film the following week. This questionnaire

made no suggestions about the content of the film, nor did it ask participants to consider any specific details from the film. The primary follow-up questionnaire asked about the five main products (salt, flour, sugar, baking soda, and baking powder) via a five-alternative forced choice (5-AFC) questionnaire (see Appendix B). The answer choices for each question included the most familiar brands of each product, though all the products used in the films were unfamiliar products. A previous study (Krug, 2004) established that a vast majority of Baylor students surveyed provided the same brand name for the products used. Therefore, the selection of the familiar brands on the follow-up questionnaire constituted “familiar false alarms.” Participants gauged their performance after answering these multiple choice questions by indicating the confidence they had in their responses on a scale from 0% (a complete guess) to 100% (absolutely certain).

Method

As participants entered the testing room, they were handed a folder containing a consent form and instructions for the experiment. Half of the participants received instructions simply asking them to pay attention to the forthcoming film and to be prepared to answer some questions about it. These instructions created an incidental encoding condition. The remaining participants received instructions explaining the purpose of the study, ones explicitly asking them to attend to the brand names of the products in an attempt to remember them. These instructions created an intentional encoding condition, one designed to confirm that the product exposures in the film were adequate for forming long-term memories. Prior to the experiment, the folders were arranged in an order alternating the two sets of instructions so that each participant

received different instructions from the one who most recently entered. The participants thus randomized themselves via their order of arrival. Participants were encouraged upon entering to address any questions they might have privately with one of the experimenters at the front of the room, so as to insure that those participants in the incidental condition did not discover the study's purpose.

When all participants had arrived and read their instructions, they were again encouraged to pay careful attention to the film and to heed their directions as they watched. The film was then shown. Afterwards, the film therapy article was distributed, and participants were simply asked to read it quietly. The participants were allowed to read for five minutes, and then they were asked to set their article aside, even if they had not completed it. Following this, members of the five minute delay group received the 5-AFC product questionnaire. Their completion of the questionnaire constituted their completion of the experiment. Members of the one week delay group received the style questionnaire (see Appendix A) following the article, and were reminded to return to the same location one week later to complete the experiment. All participants placed their completed materials back in the folder they received upon entering, and after their departure the questionnaires were sorted into the appropriate groups. When the one week delay participants returned the following week, they were asked to fill out the same 5-AFC product questionnaire completed in the five minute delay condition, and this completed their participation.

Results

Analyses. A two-by-two ANOVA test was conducted between groups to examine differences in recognition depending on the type of instructions received and the length of the delay between viewing the film and answering the primary questionnaire. The dependent measures were the number of correct responses, or hits, and the number of answers constituting familiar false alarms. On all questions subjected to analysis, participants indicated their confidence in the accuracy of their answers using a 0% to 100% scale, with choices at 20% intervals. Calibration curves plotting answers made at given confidence levels against the number of correct responses made at those levels were computed for participants in each condition. Mean gamma statistics describing the correlation between confidence and accuracy in each condition were also computed. These statistics are discussed further in a later section. All tests of statistical reliability were conducted at $p < .05$.

Accuracy measures. The primary measure of performance was overall accuracy on the follow-up recognition questionnaire. Choosing the brand name that appeared in the film constituted a hit, while choosing the most popular brand name, as established by previously acquired normative data, constituted a familiar false alarm. These results are shown in Table 1.

Hits were more frequent in the five minute delay groups than in the one week delay groups, $F(1, 371) = 44.18$, $MSE = 2.14$. Conversely, familiar false alarms were less frequent in the five minute delay groups than in the one week delay groups, $F(1, 371) = 26.38$, $MSE = 1.63$. Following the delay of one week, correct responses were

made with equal frequency as familiar false alarms. Comparable analyses of differences in hit rates between the incidental encoding groups (INC) and the intentional encoding groups (INT) were also conducted. As expected, hit rates were higher in the INT condition, $F(1, 371) = 20.07$, $MSE = 2.14$. Differences in the number of familiar false

Table 1. Means, Experiment 1

Condition	Five Minute Delay	One Week Delay
Incidental Encoding	Hit: 53.85 (4.91) FFA: 17.31 (3.72) CONF: 61.35 (3.35) GMA: .256 (.178)	Hit: 26.03 (5.17) FFA: 42.47 (5.82) CONF: 19.73 (3.20) GMA: -.101 (.261)
Intentional Encoding	Hit: 80.00 (3.92) FFA: 9.52 (2.87) CONF: 75.62 (2.94) GMA: .617 (.198)	Hit: 43.33 (5.25) FFA: 27.78 (4.74) CONF: 49.78 (3.19) GMA: -.128 (.181)

Note: Standard error of means reported in parentheses. FFA = familiar false alarm, CONF = confidence rating, GMA = gamma score

alarms in INT groups and INC groups were also statistically significant, $F(1, 371) = 7.07$, $MSE = 1.63$. Hits outnumbered familiar false alarms in both conditions, though those participants intentionally encoding the product brand names recorded more correct responses and fewer familiar false alarms than those relying on incidental encoding. The 2 X 2 ANOVA revealed that there were no significant interactions between the two independent variables. Therefore, the differences depicted in Figure 1 are the result of two substantial main effects of delay length and participant intentions.

Confidence/accuracy relationship. Loftus et al. (1989) found that participants could estimate their accuracy on eyewitness identification questions with some precision if they had not been exposed to any conflicting information. The present study replicates

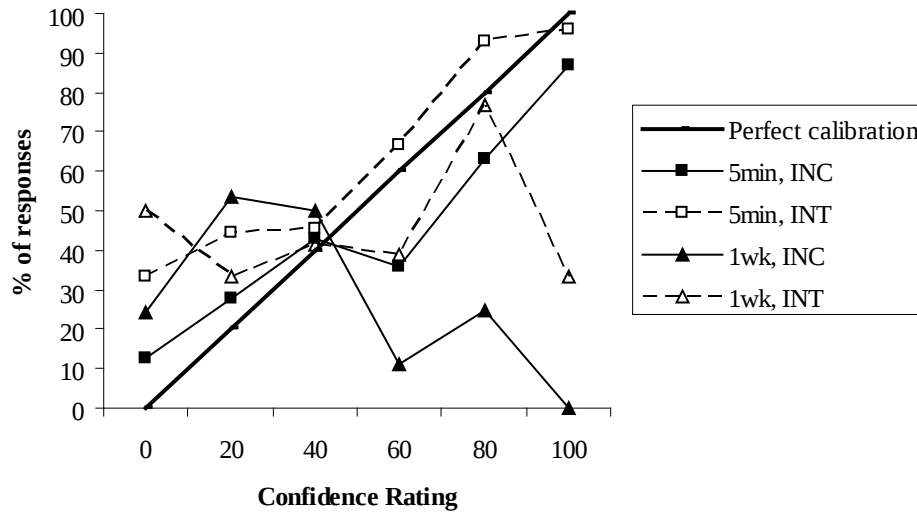


Figure 1. Confidence calibration curves for hits, Experiment 1.

these findings to a certain extent. In a plot of correct responses against confidence, perfect calibration occurs when items estimated at 100% confidence are answered correctly 100% of the time, items estimated at 80% are accurate 80% of the time, and so on. Figure 1 depicts the confidence calibration curves for each of the four conditions in Experiment 1. The area below the line of perfect calibration reflects overconfidence, while the area above the line reflects underconfidence. The curves of both one week delay groups dramatically display overconfidence, indicating that accuracy actually decreases as confidence increases.

A comparison of mean accuracy and mean confidence in each condition reveals an impressive level of calibration (see Table 1). The mean gamma statistic (G) for

participants in each condition is also displayed in Table 1¹. This statistic measures the degree to which fluctuations in accuracy are reflected by appropriate fluctuations in confidence, and vice versa. G was computed for each participant, and then averaged to a total score for each condition. Though the scores vary considerably, differences in mean G between the four conditions were not significant, $F(3, 51) = 2.63$, $MSE = 1.04$. A likely explanation for this lack of significance is the low number of participants for which G could be computed. The gamma statistic can only be calculated for participants whose accuracy and confidence actually fluctuates. If participants never answer a question correctly (or incorrectly), or never change their level of confidence—both frequent occurrences—then there is no variability to measure.

Discussion

Both the intentions of participants during encoding and the length of the delay between exposure and testing had significant effects on the frequency of both hits and familiar false alarms. The effects of intentionally encoding the brand names in the film are not surprising. The high accuracy rates at short delays suggest that the film provided adequate exposure to the products and their brand names to justify continued reliance upon it as a tool for such exposure. Somewhat surprising was the finding that informing participants precisely of what will eventually be asked of them had a less robust effect on accuracy rates than did the length of the delay.

¹ G correlations range from -1.0 (perfect negative correlation) to +1.0 (perfect positive correlation), with 0 representing complete lack of predictive accuracy. Unlike other correlation coefficients, G is not interpreted in terms of variance accounted for, but rather has a probabilistic interpretation. Specifically, if an individual gives two items different JOLs and only one of these items is correctly recalled, the probability (P) that the correct item was given a higher JOL is determined by the equation: $P = 0.5 + 0.5G$.

The 1wk, INC scenario is in many ways the most compelling, as it most closely replicates the real-world conditions under which performing such a difficult memory retrieval task would take place. Victims of asbestos-related diseases have to wait an extensive amount of time before any medical effects become evident, and during this interval there is little impetus for rehearsal. Even in the absence of conflicting post-event information, accuracy rates in the 1wk, INC condition were scarcely above chance, while familiar false alarms were twice as frequent and occurred at rates higher than those noted in the other three conditions. The low accuracy rates make reflected the difficulty inherent in such a demanding memory task, while the high false alarm rate illustrated the ease with which familiarity can be misconstrued as a specific episodic memory.

The low accuracy rates observed in the one week delay conditions were not caused exclusively by encoding failures, as the proportion of hits following the shorter delay are considerably higher. Specifically, the drop in accuracy from 67% in the five minute delay condition to 36% in the one week delay condition indicated that memories of the product brand names were encoded but decayed rapidly during the one week interval. While participants were not questioned about the extent to which they rehearsed the brand name information during the week separating the sessions, and as such no valid estimations of rehearsal can be produced, extensive attempts to retain the information still seem rather unlikely. Fortunately, this assumed element of indifference by participants mirrors the level of intent mesothelioma victims likely took in the products that might eventually cause their illness. These people likely never received specific instructions to retain such information at the time of the original event because it might later become relevant. The participants in the experiment even had the advantage

of relying on the testing room and other elements of the environment as retrieval cues, factors that have been shown to enhance retrieval in other memory paradigms via a phenomenon termed the context-dependency effect (Godden & Baddeley, 1975). Even under these most ideal conditions for forming and retrieving long-term memories, the capability for recognizing the correct brand from a short list has dropped below fifty percent following an interval of only a week.

The confidence participants placed in their memories was fairly well calibrated to their accuracy following brief delays, though even those relying on incidental encoding became overconfident after only five minutes. This decent calibration vanished in the one week delay conditions, despite the absence of any directly suggested misinformation. Though conflicting brand information was not introduced within the experiment itself, participants likely had some interaction with some of the same type of products during the one week interval between exposure and testing. The fact that confidence increased as accuracy decreased following an interval as short as one week confirms the unreliability of confidence as a predictor of accuracy. The role of juror reliance on confidence in testimony is examined in more detail in Experiment 2.

Experiment Two

One concern for those involved in product liability litigation is the potential for suggestibility during the identification process. As discussed earlier, witnesses often have their memories “refreshed” by viewing photographs prior to product identification (see Biederman et al., 1998 and Brickman, 2004 for more detailed examples). Furthermore, witnesses usually make initial product identifications in the presence of their attorneys, often having reviewed books containing dozens—even hundreds—of

possible products. These situations provide ample opportunity for suggested information to be incorporated into a witness's memory.

Experiment 2 examined the impact of post-event suggestion of brand names. These conflicting brand names were introduced in a questionnaire filled out before the 5-AFC product identification test. Because product memories were intact following a short delay but virtually absent following a one-week delay, Experiment 2 was conducted utilizing both a short, ten-minute delay and the longer one-week delay. This allowed conclusions to be drawn regarding the effects of product misinformation when original memories were intact as well as when they had faded. The ten minute delay experiment is reported as Experiment 2a, while the one week delay version is reported as Experiment 2b.

Experiment 2a

Participants

One hundred fourteen students from Baylor University participated for extra credit. Students who participated in Experiment 1 were excluded from Experiment 2.

Materials

The same film and general procedure described in Experiment 1 was used again. A table of one hundred simple math problems was used as a distractor task under the guise that the participants' performance would provide normative data for an unrelated experiment. Three versions of a similar questionnaire were used to introduce misinformation to participants. One version suggested unfamiliar, off-brand product names, another suggested the most familiar brand names, and a third control

questionnaire did not suggest any brand names. None of the questionnaires suggested brand names that actually appeared in the film. The questionnaires presented conflicting brand names without specifically asking about them, and as such exposed participants to misinformation without directly drawing their attention to it or asking about the brand name itself. The manipulation was the simple inclusion or omission of a select word or two. These critical words were not bolded, underlined, or made any more prominent than the words surrounding them in the sentence. For example: “The Martha White flour container wasn’t of normal size. Did it seem smaller or larger than normal?” exposes participants to misinformation (“Martha White”), while “The flour container wasn’t of normal size. Did it seem smaller or larger than normal?” does not. The follow-up was again a 5-AFC product test with confidence assessed at 20% intervals (see Appendix B).

Method

Once all the participants were present, they were instructed to pay careful attention to the upcoming film and prepare to answer questions regarding it. Following the film, participants were allowed to work on the math problems for five minutes to provide a buffer between the film and questioning. Participants received one of the three misinformation-providing questionnaires after the math problems. This created three misinformation groups: a control group, a familiar misinformation group, and an unfamiliar misinformation group. The three questionnaire versions were arranged in an alternating manner so that an equal number of participants received each questionnaire. Following another brief distractor, participants completed the primary recognition questionnaire.

Results

Analyses. The methods of analysis employed were similar to those used for Experiment 1. 1 X 3 ANOVAs were conducted to examine differences in accuracy, false alarm rates and confidence between members of the various conditions. The computation of Gamma was discontinued in this and all subsequent experiments in anticipation of low

Table 2. Means, Experiment 2a

Condition	Control	Familiar	Unfamiliar
Ten Minute Delay	Hit: 44.6 (3.6) Conf: 77.7	Hit: 31.6 (3.4) Conf: 80.0	Hit: 30.6 (3.4) Conf: 85.8
	FFA: 29.7 (3.3) Conf: 31.4	FFA: 59.6 (3.5) Conf: 70.8	FFA: 19.4 (2.9) Conf: 57.1
	UFA: 8.7 (2.0) Conf: 21.2	UFA: 3.1 (1.3) Conf: 46.7	UFA: 41.1 (3.7) Conf: 62.2
	CONF: 49.5	CONF: 70.8	CONF: 66.4
	O/U: 5%	O/U: 39%	O/U: 36%

Note: Standard error of means reported in parentheses. FFA = familiar false alarm, UFA = unfamiliar false alarm, CONF = overall confidence (including misses that were not designated false alarms), Conf = confidence in given type of response, O/U = over/under percentage

hit rates. Calibration curves were again constructed to examine the confidence/accuracy relationship, while over/under percentages were computed for each participant and then averaged by condition. Over/under is computed by subtracting mean accuracy from mean confidence, and is reported on a scale from -100% (complete underconfidence) to 100% (complete overconfidence). The significance level adopted for all statistical tests

was again $p < .05$. η^2 was used to measure effect size. Tukey's HSDM test was used for all post-hoc analyses in this and all subsequent experiments. The adopted significance level for these tests was also $p < .05$.

Accuracy measures. Hits and familiar false alarms were defined as above, while unfamiliar false alarms were defined as the choice of the unfamiliar product suggested in the misinformation-providing questionnaires. Selection of the suggested unfamiliar product in non-misinformation conditions was used as a baseline. Table 2 contains the mean rates of hits, familiar false alarms and unfamiliar false alarms, as well as the mean confidence of responses made in each condition.

Overall, hit rates differed between the misinformation groups, $F(2,565) = 5.18$, $SEM = 4.83$, $\eta^2 = .02$. Tukey's post-hoc analysis revealed that the difference in hits between the familiar and unfamiliar misinformation group was not significant. Considerably more familiar false alarms were committed in the familiar

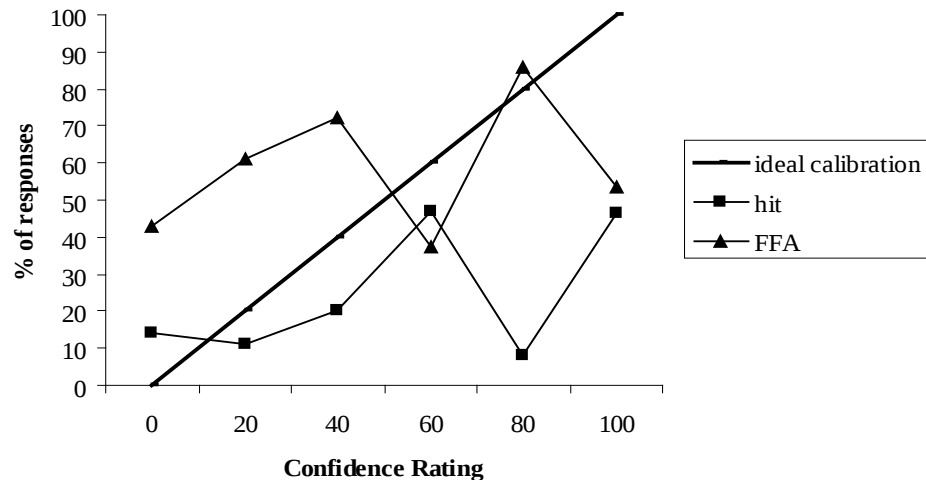


Figure 2. Confidence calibration curves in the ten-minute delay, familiar misinformation condition, Experiment 2.

misinformation group, $F(2,565) = 40.17$, $SEM = 4.59$, $\eta^2 = .12$. Post-hoc analysis revealed the difference in familiar false alarms following between the control group and the unfamiliar misinformation group was not significant. Likewise, considerably more unfamiliar false alarms were committed in the unfamiliar group, $F(2,565) = 67.57$, $SEM = 3.44$, $\eta^2 = .19$. Post-hoc analysis again confirmed the difference between control and familiar misinformation group was not statistically reliable. Mean rates of responding are reported in Table 2.

Confidence. Mean confidence ratings were computed overall and for each of the response types in each condition (see Table 2). Overall confidence in the two misinformation conditions was reliably higher than that in the control condition, $F(2,565) = 20.68$, $SEM = 1.49$, $\eta^2 = .07$.

Confidence/accuracy relationship. Mean O/U ratings were computed for each participant and averaged by condition. These means are reported in Table 2. Participants in the two misinformation conditions were considerably more overconfident than those in the control condition, $F(2, 109) = 19.65$, $SEM = 2.84$, $\eta^2 = .27$. Even at this short delay, high confidence was actually more predictive of incorrect answers than correct ones.

Calibration curves for hits and familiar false alarms were computed for familiar misinformation conditions following the ten minute delay (Figure 2)². Even following a short delay, confidence was not a reliable predictor of accuracy following exposure to familiar misinformation.

²Although presented in the same way, note that calibration curves for false alarms represents confidence plotted against false alarms, not correct responses. Thus, one should not interpret “calibration” in the same way—there is no perfect calibration when discussing false alarms.

Experiment 2b

Participants

One hundred seven Baylor undergraduate participated for extra credit. Students who had participated in Experiments 1 or 2a were excluded.

Materials

The materials were identical to those used in Experiment 2a.

Method

The same procedure was followed as in Experiment 2a, except participants completed the misinformation questionnaire and 5-AFC product identification test in a second session one week later.

Results

Analyses. The same analyses and measures of accuracy, confidence and calibration used in Experiment 2a were used again.

Accuracy measures. There were no statistically significant differences in the proportion of hits made by participants exposed to the three different misinformation-containing questionnaires, $F(2, 387) = 1.74$, $SEM = 1.99$, $\eta^2 = .01$. There was, however, a non-significant trend towards higher accuracy in the control group compared to the familiar group ($p < .26$) and the unfamiliar group ($p < .22$). Substantial differences existed in the rates of familiar false alarms committed by these three groups, $F(2, 387) = 22.12$, $SEM = 2.09$, $\eta^2 = .10$, demonstrating considerable effects of the suggestion of familiar products. Post-hoc analysis confirmed the difference between the

control and unfamiliar misinformation groups was not significant. Comparison of differences in the rates of unfamiliar false alarms between the three groups revealed an even more robust effect of misinformation, $F(2, 291) = 54.13$, $SEM = 1.13$, $\eta^2 = .27$. Again, Tukey's post-hoc analysis revealed the difference between control and familiar misinformation was not statistically reliable.

Confidence. A 1 X 3 ANOVA revealed significant differences in confidence between each of the three misinformation conditions (see Table 3). Participants exposed to familiar misinformation exhibited higher confidence ($M = 63.5$) than those exposed to unfamiliar misinformation ($M = 48.1$), who in turn exhibited higher confidence than those in control conditions, ($M = 35.7$), $F(2, 389) = 26.23$, $SEM = 3.85$, $\eta^2 = .12$.

Confidence/accuracy relationship. Participants attributed higher confidence to their familiar false alarms ($M = 71.6$) than to their hits ($M = 54.5$), $F(1, 110) = 5.20$, $SEM = 3.02$, $\eta^2 = .05$. Familiar false alarms were made with higher confidence in the familiar conditions ($M = 71.6$) than in both the control conditions ($M = 38.3$) and the unfamiliar conditions ($M = 33.3$), $F(2, 182) = 34.38$, $SEM = 2.51$, $\eta^2 = .27$. In fact, more than 75% of responses made with 80% and 100% confidence in the familiar condition were familiar false alarms. More familiar false alarms were made than hits at each of the six confidence ratings following familiar misinformation exposure, indicating that regardless of the confidence rating utilized, identifications were more likely to be familiar false alarms than correct answers. O/U percentages reveal overconfidence in each misinformation condition, the most robust resulting from familiar suggestions. Both familiar misinformation and unfamiliar misinformation induced greater overconfidence

($M = .47$ and $M = .32$, respectively) than that noted in control conditions ($M = .11$), $F(2, 95) = 11.03$, $SEM = .04$, $\eta^2 = .19$.

Discussion

The presentation of inaccurate post-event information had a profound effect on the ability to recognize product brand names from a previous exposure following both

Table 3. *Means, Experiment 2b*

Condition	Control	Familiar	Unfamiliar
One Week	Hit: 24.4 (3.8) Conf: 50.3	Hit: 16.8 (3.3) Conf: 54.5	Hit: 16.3 (3.2) Conf: 51.8
	FFA: 41.7 (4.4) Conf: 38.3	FFA: 68.7 (4.1) Conf: 71.6	FFA: 31.1 (4.0) Conf: 33.3
	UFA: 7.3 (2.7) Conf: 14.3	UFA: 2.0 (1.4) Conf: 0	UFA: 47.1 (4.9) Conf: 60.9
Delay	CONF: 35.7	CONF: 63.5	CONF: 48.1
	O/U: 11%	O/U: 47%	O/U: 32%

Note: Standard error of means reported in parentheses. FFA = familiar false alarm, UFA = unfamiliar false alarm, CONF = overall confidence (including misses that were not designated false alarms), Conf = confidence in given type of response, O/U = over/under percentage

short and long delays. This finding is especially intriguing in light of the subtlety with which the misinformation was presented. Such a gentle, inconspicuous presentation of information stands in stark contrast to the more direct manner in which some attorneys

“refresh the recollections” of witnesses by showing them photographs or lists of products (Biederman et al., 1998; Brickman, 2004).

Experiment 1 revealed a decent ability to remember products at a short delay, but this decayed rapidly over a week’s time. Indeed, accuracy at these long delays is sufficiently poor that our post-event suggestions may be competing with few actual memories. More likely, these suggestions are used to fill in voids in recollections. The findings of subsequent experiments described below support this conclusion.

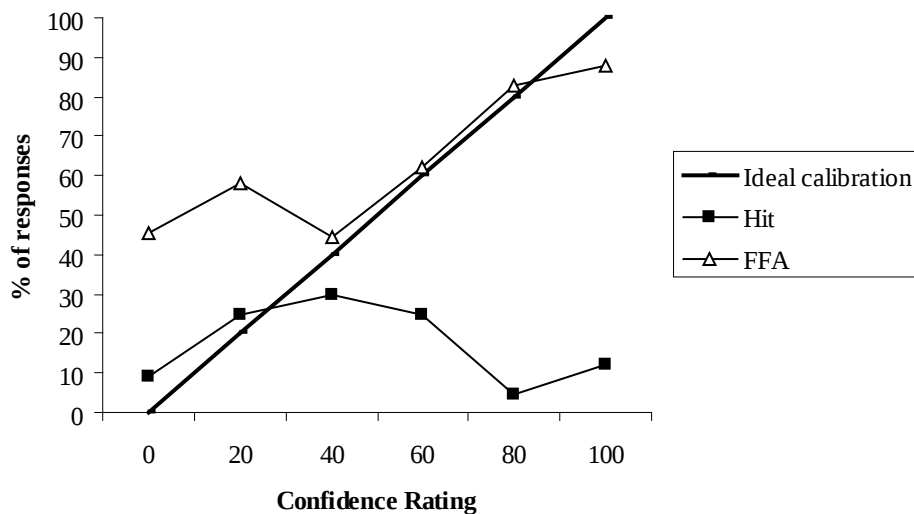


Figure 3. Confidence calibration curves in the one-week delay, familiar misinformation condition, Experiment 2.

Substantial misinformation effects were seen following both short and long delays in Experiment 2. Following only ten minutes, the suggestions likely eradicate the already fading memories of the film, while after a week the effect likely results from a combination of familiarity bias and outright false memory creation. The bottom line is that participants had very little memory left to distort, and the basis of their responding was provided almost entirely by the experimenter via post-event suggestions.

Regardless of the underlying mechanism leading to the decision, a statistically reliable increase in confident but inaccurate responding was seen following exposure to both familiar and unfamiliar suggestions.

The drastic increase in confidence placed in both familiar false alarms and unfamiliar false alarms when that type of misinformation has been suggested was striking. Misinformation effects were especially profound for brand names that were already familiar to participants. This familiarity bias was first noted by Krug and Weaver (2005) and illustrates the indirect manner in which product identifications are often made: participants appear to mistake a sense of familiarity for true recollection. The confidence ratings depicted in Tables 2 and 3 and in Figures 2 and 3 support this conclusion. While it is possible that those participants who did not remember the products from the film chose those products most familiar to them as logical guesses, their confidence ratings should have reflected this strategy. In contrast, familiar misinformation led to the highest overall confidence ratings, as well as to higher confidence in familiar false alarms than in hits. Across the three misinformation conditions in Experiment 2b, answers made with 100% certainty were three times more likely to be familiar false alarms than hits.

These distortions in witness confidence are important because of their interpretation by jury members. Koegh and Markham (1998) concluded that the primary criteria used by jurors to determine the believability of witnesses are the number of sensory details used in their descriptions and the confidence they place in their testimony. Numerous other studies confirm that jurors rely most heavily on the confidence exhibited by witnesses when attempting to assess their credibility (Lindsay, Wells, & Rumpel, 1981; Penrod & Cutler, 1995; Wells, 1993). The high confidence placed in false alarms

is concerning, as following a discovery session in which numerous photos of potential products were viewed, plaintiffs in product identification cases could be quite confident in their prior exposure to certain products.

The present experiment created substantial misinformation effects with minimal deception and suggestiveness. The induction of considerable changes in responding with such an insignificant exposure to a brand name seems analogous to what is described in the literature as a mere exposure effect (Bornstein & D'Agostino, 1992; Gordon & Holyoak, 1983). Zajonc (1968) originally described the mere exposure effect as the observation that repeated, unreinforced exposure was sufficient to alter attitude towards a stimulus. The phenomenon is typically regarded as an artifact of implicit memory, as exceptionally brief exposures too fleeting to enter into conscious awareness still have reliable effects on later behavior. Other more conspicuous tests of implicit memory have been conducted, some of the most notable by Jacoby (see Jacoby & Dallas, 1981; Jacoby, Toth & Yonelinas, 1993). In these experiments, participants may be unable to identify previously studied words (either by free recalling them or recognizing them in another list), but still exhibit enhanced performance on unrelated tests involving these words for which they have no conscious recollection of studying. Their mere exposure to the word “primes” them for improved performance on fragment completion tests and speeded perceptual identification tests when compared to words which have not been primed with implicit exposures.

While no tests were administered in the present study to ascertain the participants' conscious awareness of misinformation, as these would have compromised the broader intent of the study, participants likely would have been unable to produce the

brand names from memory, given the limited and insignificant exposure. Therefore, the increased rate of selecting the presented misinformation during the recognition follow-up test could be interpreted as the mere exposure to the brand names in the previous tests facilitating the choice of those names. The exceptionally high rate of familiar false alarms in the familiar misinformation group could reflect the combination of mere exposure effects with the gravitation towards the most salient brand name observed in Experiment 1. Those researchers explaining the intrusion of conflicting post-event information during retrieval as problems of source attribution (Chandler et al., 2001; Lindsay, 1990) claim that individuals can remember both the original event and the post-event suggestion, but simply cannot distinguish between the two. This explanation does not appear tenable for such long-term memories of insignificant events.

Regardless of the theoretical explanations, it is clear that post-event suggestions—even subtle ones—had a dramatic impact on recollections of insignificant parts of previous experiences. The subtle presentation of misinformation also greatly influences the confidence individuals had in their recollection, an interesting consequence given the decrease in accuracy and the dramatic rise in false alarms produced by such exposure.

Experiment Three

Experiment 3 refreshed participants with photos of the products they might have seen, and differs from Experiment 2 in that participants were aware of the suggestions they were receiving. The experimental paradigm employed here is much more similar to that used by attorneys, but differs in that participants only saw two brands of each product, rather than tens or even hundreds of possible choices. Like Experiment 2, two different versions of Experiment 3 were conducted. Experiment 3a refreshed participants

following a ten minute delay, while Experiment 3b refreshed them following a one week delay.

Experiment 3a

Participants

One hundred twenty Baylor undergraduates participated for extra credit. Those participating in the first two experiments were excluded.

Materials

The same film and recognition questionnaires used in Experiment 2 were used again. The 40-item vocabulary portion of the Shipley Institute of Living Scale (see Zachary, 1986) and the style questionnaire from Experiment 1 were used as five-minute distractor tasks. Color photos of products were presented via Microsoft Power Point onto a projection screen at the front of the testing room.

Method

The general methodology was similar to that of Experiment 2. Participants were tested in groups of around 30. Each group was told to pay careful attention to the film and to be prepared to answer questions regarding it. Following the film, participants completed the Shipley vocabulary test and the style questionnaire. Completion of these two questionnaires established a ten minute delay between exposure and testing. Participants were then shown photo slideshows of different brands of products. After explaining to participants that they would be questioned on the brand names they had seen previously, each slideshow began with these instructions: “Before attempting to remember the brands, watch the following slideshow of products you may have seen.

Use these pictures to jog your memory—you will not be allowed to view them while you fill out the questionnaire.” The show contained six slides, each shown for eight seconds and containing two brands of the same product side-by-side. There were three photo combinations: Correct/familiar, correct/unfamiliar, and familiar/unfamiliar. Each participant saw two slides of each combination, and three slideshows were made to insure that all 18 products were equally suggested across conditions. The photo conditions were thus completely counter-balanced across participants. After the slideshow, participants completed the same recognition questionnaire used in Experiment 2.

Results

Analyses. The same measures of accuracy and confidence and the same methods of analysis used in Experiment 2 were used again. 1 X 3 ANOVAs were conducted between conditions for both the ten-minute and one-week delay studies. Means are reported in Table 3.

Accuracy measures. Hit rates were high and nearly identical in conditions viewing the correct photo, but drastically lower when two incorrect photos were seen, $F(2, 591) = 160.8$, $SEM = 3.66$, $\eta^2 = .30$. Both familiar false alarms-- $F(2, 849) = 28.11$, $SEM = 2.87$, $\eta^2 = .06$ and unfamiliar false alarms-- $F(2, 849) = 66.57$, $SEM = 3.24$, $\eta^2 = .14$ --occurred more frequently when two incorrect photos had been viewed. At this short delay, correct responding dominated conditions where the previously seen product was contained in the slideshow.

Confidence. Mean overall confidence as well as confidence in each type of response is reported in Table 4. Mean confidence in the familiar/unfamiliar condition was lower than in conditions presenting the correct photo, $F(2, 845) = 7.02$, $SEM = 1.78$, $\eta^2 = .02$. Tukey's post-hoc analysis revealed no other significant differences.

Confidence/accuracy relationship. Confidence was higher in hits than in misses across all conditions, $F(1, 766) = 122.51$, $SEM = 1.08$, $\eta^2 = .14$. Photo refreshing did not significantly affect confidence placed in hits or unfamiliar false alarms, though familiar/unfamiliar refreshing did significantly boost confidence placed in familiar false alarms, $F(2, 120) = 13.36$, $SEM = 2.87$, $\eta^2 = .12$. A 1 X 3 ANOVA in O/U revealed that

Table 4. Means, Experiment 3a

Condition	Corr/Fam	Corr/Unfam	Fam/Unfam
Ten Minute Delay	Hit: 83.0 (2.5) Conf: 87.6	Hit: 83.2(2/4) Conf: 88.5	Hit: 23.4 (2.8) Conf: 85.2
	FFA: 13.9 (2.3) Conf: 47.5	FFA: 3.0 (1.1) Conf: 37.1	FFA: 27.3 (2.9) Conf: 69.0
	UFA: 2.6 (1.1) Conf: 43.3	UFA: 12.1 (2.1) Conf: 55.0	UFA: 46.3 (3.3) Conf: 65.8
	CONF: 80.4	CONF: 82.4	CONF: 69.7
	O/U: 3%	O/U: -1%	O/U: 46%

Note: Standard error of means reported in parentheses. Corr = correct photo, Fam = familiar photo, Unfam = unfamiliar photo, FFA = familiar false alarm, UFA = unfamiliar false alarm, CONF = overall confidence (including misses that were not designated false alarms), Conf = confidence in given type of response, O/U = over/under percentage

participants were much more overconfident following familiar/unfamiliar refreshing, $F(2, 755) = 106.04$, $SEM = 1.72$, $\eta^2 = .22$. Post-hoc analysis revealed no other significant differences. Participants were reasonably-well calibrated when viewing the correct product after only a ten minute delay.

Calibration curves were computed, but are not presented here. These were largely similar to the curves noted in Experiment 3b below, and thus discussion of the effects of photo refreshing on calibration is reserved for a later section.

Experiment 3b

Participants

One hundred twenty Baylor undergraduates participated for extra credits. Those who had participated in any previous experiments in this study were excluded.

Materials

The materials were identical to the ones used in Experiment 3a.

Method

As in Experiment 2, the only facet of methodology changed in the second experiment was the delay between exposure and refreshing. Participant completed all distractor questionnaires in the first session, and began the second session one week later by viewing the refreshing array.

Results

Analyses. The same analyses and measures of accuracy and confidence described in Experiment 3a were used again.

Accuracy measures. Hit rates in the three photo conditions were all reliably different, $F(2, 591) = 65.28$, $SEM = 4.30$, $\eta^2 = .18$, with the most prominent difference being the lack of hits when two incorrect photos were viewed. Similar differences existed in the rates of familiar false alarms, $F(2, 591) = 54.78$, $SEM = 4.45$, $\eta^2 = .16$. Unfamiliar false alarms were reliably more frequent when unfamiliar photos had been viewed, $F(2, 591) = 6.28$, $SEM = 4.11$, $\eta^2 = .02$.

Table 5. *Means, Experiment 3b*

Condition	Corr/Fam	Corr/Unfam	Fam/Unfam
One Week	Hit: 40.9 (3.5) Conf: 76.5	Hit: 53.5 (3.6) Conf: 74.0	Hit: 6.1 (1.7) Conf: 76.7
	FFA: 38.4 (3.5) Conf: 47.9	FFA: 12.1 (2.3) Conf: 47.5	FFA: 58.6 (3.5) Conf: 49.8
	UFA: 13.6 (2.4) Conf: 34.8	UFA: 23.7 (3.0) Conf: 44.3	UFA: 27.8 (3.2) Conf: 56.0
	CONF: 56.7	CONF: 59.9	CONF: 52.5
	O/U: 15%	O/U: 6%	O/U: 53%

Note: Standard error of means reported in parentheses. Corr = correct photo, Fam = familiar photo, Unfam = unfamiliar photo, FFA = familiar false alarm, UFA = unfamiliar false alarm, CONF = overall confidence (including misses that were not designated false alarms), Conf = confidence in given type of response, O/U = over/under percentage

Confidence. Though overall confidence dropped slightly in the familiar/unfamiliar photo condition following a one-week delay (see Table 5), the differences in overall confidence between the three photo combinations following this longer delay were not statistically reliable, $F(2, 591) = 2.21$, $SEM = 2.51$, $p < 0.11$, $\eta^2 =$

.01. This is striking, given the substantial decrease in accuracy in the familiar/unfamiliar condition.

Confidence/accuracy relationship. Confidence was higher in hits than misses across conditions following the one week delay, $F(1, 493) = 83.13$, $SEM = 1.55$, $\eta^2 = .14$. While the refreshing conditions did not affect confidence in hits or familiar false alarms, unfamiliar false alarms were made with higher confidence following familiar/unfamiliar refreshing, $F(2, 126) = 4.36$, $SEM = 2.86$, $\eta^2 = .07$. A 1 X 3 analysis of O/U revealed that participants were considerably more overconfident following refreshing with exclusively incorrect products, $F(2, 492) = 51.96$, $SEM = 2.16$, $\eta^2 = .17$. Post-hoc analysis revealed no other differences were significant.

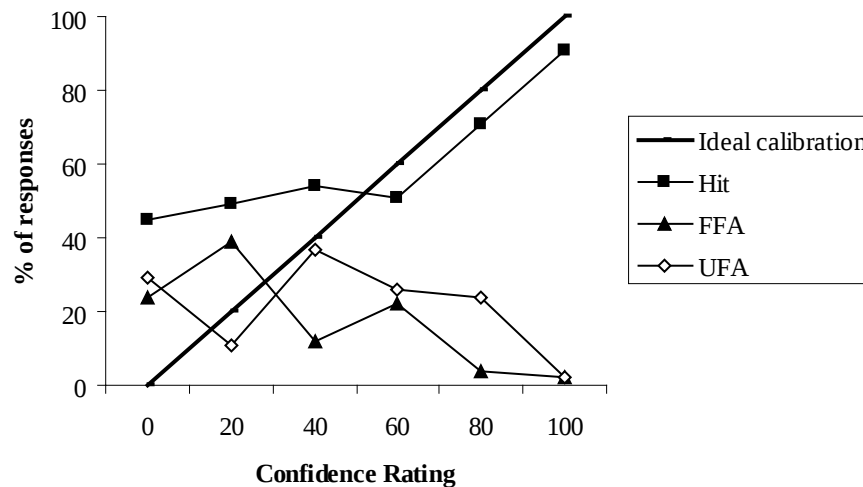


Figure 4. Confidence calibration curves following refreshing with the correct product, Experiment 3b.

For the computation of confidence calibration curves, the refreshing conditions containing the correct product were merged together—these curves are shown in Figure 4, while curves for the familiar/unfamiliar refreshing condition are shown in Figure 5.

Even following a one week delay, decent calibration emerged when the correct photo had been viewed. However, when two incorrect photos were viewed following the one week interval, the curves revealed poor calibration similar to that noted in the familiar misinformation condition of Experiment 2. As in Experiment 2, this overconfidence results not from the inflation of confidence, but rather from the prominent deflation in accuracy following inaccurate post-event suggestions. As mentioned above, a similar pattern of calibration was noted following the ten minute delay in Experiment 3a.

Discussion

Experiment 3 demonstrated the powerful effects of photo refreshing in a product identification situation. Following a short delay, the presentation of the correct product

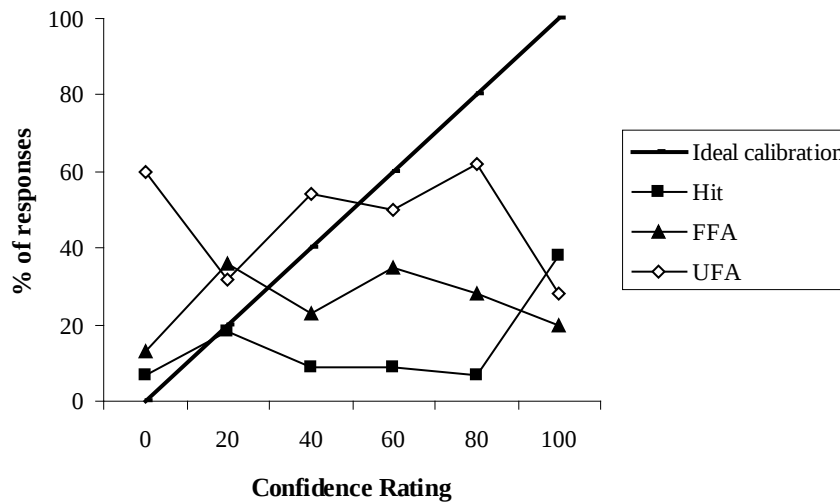


Figure 5. Confidence calibration curves following refreshing with exclusively incorrect products, Experiment 3b.

boosted accurate responding to over 80%, and following a one-week delay increased accuracy to nearly twice the baseline rates observed in Experiment 1. The ability to answer correctly following refreshing with the true product suggests that some episodic

trace of the product's appearance in the film is likely stored in memory. However, true memory "jogging" appears only to occur when the correct product is viewed, as refreshing arrays containing exclusively incorrect photos profoundly hindered correct identification. Even at a short delay, the presentation of two incorrect photos dropped accuracy to scarcely above chance, while at the more relevant one-week delay, accuracy plummeted to an abysmal 6%. Across the two experiments, the *lowest* rate of choosing one of the two viewed photos at follow-up was 74%--this after viewing two incorrect photos at a short delay. When viewing two incorrect photos after a week's delay, participants erred in this fashion a striking 86% of the time.

If one assumes that the exclusive effect of photo refreshing is to facilitate access to existing memories, the biasing effects of familiar/unfamiliar refreshing is puzzling. Such an explanation would account for the elevated hit rate, but not for the inflated false alarm rates in conditions when the correct photo was not shown. In contrast, it appears that photographs operate as a source of suggestion—not as facilitation of retrieval of existing memories. Photos appear to be acting as perceptually-rich, highly persuasive post-event suggestions, profoundly skewing responding toward the two products presented. While it is possible that correct photos were refreshing rapidly-decaying—but still present—memory traces following the short delay, it seems less likely that increases in accuracy following a week resulted from actual memory jogging. Experiment 1 supported the notion that product information still lingered following short delays, but that a one week delay was sufficient to eradicate most product memories, leaving little to refresh. The near-ceiling hit rates observed following short delays and the presentation of the correct photo likely resulted from post-event suggestions

confirming memories. Following a week long delay, the strength of the memory trace had likely faded considerably, and only accessed a vague sense of familiarity with the correct product relative to the distractor.

The familiar/unfamiliar conditions revealed that—even following only ten minutes—what remained of product memories was usually insufficient to resist the influence of the refreshed products. Even when memories were most accessible, participants were overwhelmed by the suggestive nature of photo refreshing. In this case, the experimenter-provided suggestions were simply better-trusted than rapidly fading episodic traces.

The confidence data indicated that following photo refreshing, most participants distinguished between correct identifications and those based on inaccurate post-event suggestions. Unlike in Experiment 2, participants placed more confidence in hits than false alarms, even following refreshing with two incorrect products. This suggests that participants only resorted to choosing from two incorrect alternatives when they could not retrieve the brand information from their memories. Unfortunately, it appears that participants rarely remembered the correct product without prompting. This scenario—which produced the lowest accuracy rate of all five experiments in this study—most closely mimics real-world product identification situations described by Biederman et al. (1998) and Brickman (2004).

Several potential processes could have promoted the inflated accuracy rates noted in correct/familiar and correct/unfamiliar refreshing conditions in Experiment 3. Participants could have remembered the products prior to refreshing and the photos simply acted to confirm their original memory trace. Similarly, correct photos could also

be allowing access to memory traces that are otherwise too weak to be recalled. Finally, participants who have no episodic memory of the original product may be adopting the correct suggestion as their memory. In this sense, correct identifications would be generated by the same process that produces misinformation effects—participants may adopt the proper suggestion simply by guessing or relying on a weak sense of familiarity sponsored by having viewed the product previously to incline them towards the correct identification.

Whatever the cause of the correct identifications following the viewing of refreshing arrays containing the correct product, Experiment 3 can essentially be summarized with a single statement: If you show participants the correct product, they are more likely to identify it; if you show participants two incorrect products, they will likely identify one of those. It is tempting to advance the explanation that in correct/familiar and correct/unfamiliar conditions, the true product provided the opportunity for memory jogging, while the unseen distractor offered a concurrent incidence of misinformation exposure. In the ten minute delay experiment, misinformation was sufficient to challenge the original memory only when the true product was not equally refreshed. The findings in the one week delay experiment are somewhat more complicated, given that accuracy in Experiment 1 was so low. This would imply that increased hit rates could not be the result of memory jogging, but rather come about via the same mechanism producing misinformation effects in other conditions.

Experiment Four

In Experiment 3, photo refreshing had profound effects on product identification. Rates of choosing one of the two refreshed products, regardless of accuracy, was unexpectedly high. Experiment 4 was designed to examine what made these photographs so suggestive, and to what extent they could distort responding.

Participants

One hundred eight Baylor University undergraduates participated for extra credit credit. Participants who had participated in Experiments 1, 2, or 3 were excluded.

Materials

Experiment 4 utilized a new film that resembled a news report on students using various condiment products. This film allowed for longer viewing times of the products on screen (around thirty seconds average), but never showed close-up images of the product containers. A new 5-AFC product identification was created (see Appendix C) which closely resembled those used in previous experiments. A new survey was also created for the new film (see Appendix D). This and the Shipley vocabulary test were again used as distractor tasks. Several new PowerPoint slideshows were created for product refreshing. These slideshows contained four types of critical photographs, three of which were achieved via PhotoShop manipulation. For some photographs, the correct brand name was superimposed onto the incorrect container. For others, the incorrect brand name was superimposed on the correct container. Other photographs combined unseen brands and unseen containers, creating products that are non-existent in the real world. Same brand/different container photographs took the brand name of a product

appearing in the film and super-imposed it on a container not appearing in the film. Same container/different brand photographs placed a brand name not seen in the film on a container that did appear in the film. Different brand/different container photographs placed a brand not seen in the film on a container not seen in the film. These manipulations created products entirely non-existent in the real world, such as Hillshire Farms ketchup, Smuckers barbecue sauce, and Oscar Mayer relish. Finally, non-manipulated photographs of the products appearing in the film were included to comprise a “true” refreshing condition. Each slide contained one of the four genres of photographs described above, as well as a distractor photograph of an unfamiliar product not appearing in the film. Participants thus saw two photographs of each product type appearing in the film. Therefore, each participant was exposed to each type of critical lure (the manipulated photo, or unaltered photo in true conditions) at least once, and sufficient slideshows were created to counterbalance the order and frequency of each lure's appearance.

Method

The methodology of Experiment 4 was nearly identical to that of Experiment 3, with the exception that only a one week delay was used. Participants viewed the film and completed the Shipley vocabulary test and the relevant style questionnaire during the first session. One week later, they were refreshed with one of the slideshows described above, and then given the 5-AFC product identification questionnaire. Several sessions were run to counterbalance the presentation of the different refreshing slideshows.

Results

Analyses. 1 X 4 ANOVAs were conducted analyzing the rates of hits, lures, and distractors in each condition. Hits were defined as above, while “lures” were defined as the selection of the critically manipulated product in the refreshing array. “Distractors” were defined as the selection of the product which appeared along with the lures in the refreshing array. Analysis of these data was less-straight forward than in previous experiments, primarily because responses driven by “same brand/different container” refreshing and those prompted by actual memory for the film exposure were indistinguishable given that ultimate identifications were made via brand name. As such, certain hypothesis-driven post-hoc comparisons were made. The assumption driving this experiment was that such difficult product identification tasks were perceptually-driven, and as such that “same container/different brand” refreshing would significantly increase identification of the super-imposed brand on the 5-AFC test. Concurrently, it was expected that “same brand/different container” refreshing would not lead to a reliable increase in accuracy, as the brand name alone would not increase brand memory in the absence of perceptually matching stimuli. In addition, similar 1X4 ANOVAs analyzing mean confidence and O/U were conducted, as were calibration curves examining the different brand/different container refreshing condition. The significance value adopted was again $p < .05$, while η^2 was again used to report effect sizes.

Accuracy measures. A 1 X 4 ANOVA revealed significant differences in hit rates, $F(3, 642) = 41.60$, $SEM = 1.84$, $\eta^2 = .163$. Tukey’s post-hoc analysis revealed that all pairwise differences between groups were significant except between

Table 6. Means, Experiment 4

Condition	Hit	Lure	Distractor	Confidence
SB/DC	52.56 (3.41)	52.56 (3.41)	29.69 (3.31)	58.79 (2.14)
SC/DB	16.00 (3.00)	29.33 (3.73)	36.67 (3.94)	53.07 (2.44)
DB/DC	6.82 (2.20)	25.76 (3.82)	56.82 (4.33)	41.39 (2.40)
True	42.95 (4.06)	42.95 (4.06)	34.23 (3.90)	52.08 (2.39)

Note: Standard error of means reported in parentheses. SB/DC = same brand/different container, SC/DB = same container/different brand, DB/DC = different brand/different container

same brand/different container and true refreshing, $p < .07$. In contrast to expectations, same brand/different container refreshing led to the highest accuracy of all conditions, though the difference when compared to accuracy following true refreshing was not statistically reliable.

Overall, the lure items were identified at different rates between groups, $F(3, 642) = 10.41$, $SEM = 1.92$, $\eta^2 = .05$. Tukey's post hoc analysis revealed that the significant differences were between same brand/different container refreshing and both same container/different brand and different brand/different container refreshing. Different brand/different container and true refreshing also led to significantly different rates of selecting the lure product ($p < .014$). The selection of same container/different brand lures was not significantly greater than the selection of different container/different brand lures, $F(1, 280) = .446$, $SEM = 2.67$, $\eta^2 = .002$.

The rate of identifying the distractor item in the different brand/different container condition was also greater than all others, $F(3, 642) = 9.14$, $SEM = 1.94$, $\eta^2 = .04$.

Post-hoc analysis revealed none of the other differences between groups were statistically reliable.

Confidence. Participants were reliably less confident in the different container/different brand condition than all others, $F(3, 642) = 9.34$, $SEM = 1.19$, $\eta^2 = .042$. Tukey's post-hoc analysis revealed no significant differences between the other three conditions—mean ratings were between 52 and 59%.

Confidence/accuracy relationship. Across the experiment, confidence was higher in hits ($M = 61\%$) than in misses ($M = 48\%$), $F(1, 644) = 27.11$, $SEM = 1.19$, $\eta^2 = .04$. Confidence in hits was reliably higher following same brand/different container refreshing than following true refreshing, $F(1, 260) = 15.02$, $SEM = 1.88$, $\eta^2 = .06$. A 1 X 4 ANOVA and subsequent Tukey's post hoc analysis of O/U reveals that participants in the same brand/different container and true refreshing conditions were less overconfident than those not viewing the correct brand, $F(3, 642) = 18.41$, $SEM = 1.99$, $\eta^2 = .08$.

Numerous calibration curves can be created from these data, though perhaps the most interesting depicts confidence in the selection of non-existent products and the distractors that accompany them in the different brand/different container condition (see Figure 6). Following refreshing with an unfamiliar product and a non-existent product, confidence in subsequent identification of the unfamiliar product skyrockets. This suggests that confidence in an identification is influenced not only by the photo ultimately identified, but also by the other photos present in the array. That is, confidence cannot be regarded as a meaningful measure of “memory strength,” because

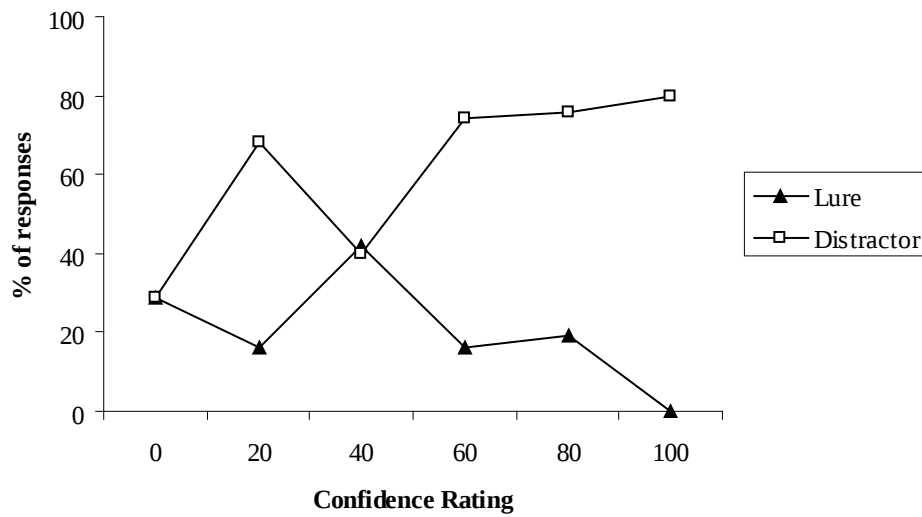


Figure 6. Confidence calibration curves, different brand/different container condition, Experiment 4.

such a judgment would depend only on the identified product, regardless of the context in which the photo was shown. Confidence was affected by the surrounding photographs, however, suggesting it is influenced by the context of retrieval, not purely the salience or strength of memory traces.

Discussion

The findings of Experiment 4 fail to support the hypothesis of perceptually-driven product identification. If selections were driven primarily by the perceptual cues available in photos used for refreshing, two findings should have emerged. First, superimposing correct brand names on different containers should have reduced their rate of selection, since the perceptual cues would have been different from those noted during exposure. Second, incorrect brand names superimposed on correct containers should have lead to high false alarm rates, as this condition preserved the original perceptual cues. Neither of these trends emerged. Instead, photos containing the correct brand

information produced the greatest accuracy—regardless of the container it was presented on. In fact, accuracy was highest following refreshing with "same brand/different container" photographs, indicating that viewing the correct brand name amidst a different set of perceptual cues benefited subsequent identification rather than hindered it. Even though the difference between conditions was non-significant, it suggests a trend opposite of the one hypothesized. Likewise, the expected increase in identifications of the incorrect brand names superimposed on the correct containers was not noted, suggesting that participants did not simply find the container that appeared most familiar and select whatever name happened to be printed on its label.

The most surprising result of Experiment 4 was the selection of non-existent products over one quarter of the time following "different brand/different container" refreshing. Unlike all false alarms in previous experiments, selection of these non-existent products could not have been the result of source monitoring errors--participants had never seen these products outside of the experiment's photo refreshing segment, and therefore their selection had no basis on witness' memories of the original event. These false alarms were indisputably the product of suggestion, and argue strongly that a similar mechanism operated in previous experiments as well.

However, consider also the tendency to identify confidently the unfamiliar distractor paired with these impossible products (see Figure 6). These distractors were selected a striking 57% of the time, making these products—ones sharing no resemblance to those actually seen—the most commonly selected response across all conditions of Experiment 4. This also marked the highest rate of unfamiliar false alarms seen in the entire study, higher even than that achieved following exposure to embedded unfamiliar

suggestions in Experiment 2. Selections of the distractor were also made with higher confidence than were correct identifications. Though this difference was not significant, it sponsors speculation that, rather than excluding these two unseen products from consideration, participants utilized two-product refreshing to make the identification task a “one or the other” decision. Apparently, participants who were able to reject the implausible products readily turned to the other alternative provided in the refreshing array. Recall that participants had been instructed that the product slideshow contained photos of products they *may* have seen in the film, and that they would eventually make their choice from five options, not exclusively the two seen. Despite this, the photographs readily served as powerfully influential post-event suggestions when participants possessed weak (or non-existent) memories of the original exposure. Correct identification occurred a scant 7% of the time in the different brand/different container condition, replicating the finding of Experiment 3 that viewing incorrect options does not make the correct option more plausible—indeed, just the opposite is true.

Experiment 4 supports the conclusion that two-product photo refreshing provides highly influential post-event suggestions rather than an objective opportunity for memory jogging. The extent to which presenting photos in an entirely neutral manner might serve as effective retrieval cues is explored in Experiment 5.

Experiment Five

Experiments 3 and 4 demonstrated that photo refreshing utilizing a sub-set of eventual answer choices had a powerful influence on subsequent identification—even when accompanied by instructions that the pictured products may not be correct. These findings supported two theories. A “refreshing” hypothesis suggests that photos augment

identifications by jogging existing product memories. Conversely, a “suggestibility” hypothesis argues that photo refreshing influences the identification process by providing suggestions that are either adopted in the absence of original memories or better-trusted than the weak memories that do exist. In Experiment 5, participants saw the same films as before, but prior to being tested were refreshed with photos of all products appearing on the ultimate identification test. In this scenario, the refreshing hypothesis would predict that viewing the correct photo should enhance a pre-existing (likely weak) memory trace of the original event, and thus should improve performance relative to a condition where no photos were shown. In contrast, the suggestibility hypothesis would predict no increase in accuracy, since presenting all possible products removes suggestibility. Experiment 5 was designed to test these competing hypotheses by comparing witnesses refreshed with “neutral” photograph arrays to those not refreshed prior to identification.

Participants

One hundred twenty one Baylor undergraduates participated for course credit or extra credit. Students participating in any previous versions of the experiment were not allowed to participate again.

Materials and Method

A similar methodology was employed to that used in the one week delay experiments described above. Both the film used in Experiments 1-3 and the film used in Experiment 4 were used to present products to participants. The difference in these films with regards to product exposure was treated as an independent variable. The cooking

show film used in the first three experiments offered brief, close-up exposures of products, while the alternate film used in Experiment 4 provided longer, more natural exposures that did not involve close-ups filling the entire screen. This latter film may best represent real-world exposures, as the products were never truly the focal point, but are shown for long periods of time and can be clearly noted if viewers take it upon themselves to do so. The cooking show is somewhat less natural, but it affords viewers the opportunity to focus briefly on little else but the relevant products, and offers less incidental exposure while other events are taking place. These two films are thus considered to provide “close-up” and “extended” exposures.

The primary factor of interest in Experiment 5 was the presence or absence of photo refreshing. Half of the participants viewed PowerPoint slideshows concurrently showing photographs of each of the five products that would be options on the 5-AFC identification test. The other half did not view photos. The order and positioning of these photographs was randomized between sessions to eliminate the possibility of biasing via the sequencing of the photographs on the slides. Six slides were created, one for each product type appearing in each of the films. Each slide was viewed for ten seconds, and showed photographs of the five brands of products side-by-side. Participants received the same instructions before viewing the slideshow as in previous experiments, and as such were told that the slides contained pictures of products they may have seen in the film. Participants again completed the 40-item Shipley vocabulary questionnaire and the appropriate film style questionnaire during the first session, and returned one week later to complete the 5-AFC product identification questionnaire. Half the participants were neutrally refreshed at the follow-up session, while half received no

refreshing whatsoever. Half viewed the close-up exposure film, while half viewed the natural exposure film, resulting in a 2X2 design crossing the opportunity for refreshing and the type of film viewed.

Results

Analyses. The primary variable of interest was accuracy following the two refreshing conditions. A 2X2 ANOVA was conducted analyzing the effects of the different methods of exposure and the presence of neutral refreshing. Mean confidence and over/under percentages were also analyzed. As in all previous experiments, the level of significance adopted was $p < .05$, with η^2 used to report effect sizes.

Accuracy measures. Mean hit rates are shown in Table 7. The differences in accuracy following close-up and extended exposure were not significant, $F(1, 722) = 0.29$, $p < .588$, $SEM = 1.63$, $\eta^2 = .00$. Neutral refreshing led to slightly higher hit rates ($M = 30\%$) than seen in the non-refreshed control groups ($M = 21\%$), $F(1, 722) = 6.80$, $SEM = 1.67$, $\eta^2 = .01$. There was no interaction between type of refreshing and type of exposure, $F(1, 722) = .332$, $p < .565$, $SEM = 1.63$, $\eta^2 = .000$. If hit rates for products in the extended exposure film are analyzed exclusively, the impact of neutral refreshing on accuracy becomes non-significant, $F(1, 322) = 1.98$, $SEM = 2.39$, $p < .16$, $\eta^2 = .01$.

Confidence. The mean confidence in each condition is shown in Table 7. Exposure to neutral refreshing did not reliably increase participant confidence, $F(1, 724) = 2.05$, $SEM = 1.24$, $\eta^2 = .00$. There was also no difference in confidence depending on

Table 7. Means, Experiment 5

Condition	Neutral Refreshing	No Refreshing
Extended Exposure	Hit: 28.08 (4.16) CONF: 45.20 (2.58) O/U: 17%	Hit: 20.29 (3.12) CONF: 39.66 (2.42) O/U: 19%
Close-up Exposure	Hit: 32.55 (3.25) CONF: 41.38 (2.26) O/U: 9%	Hit: 20.59 (3.23) CONF: 38.55 (2.65) O/U: 18%

Note: Standard error of means reported in parentheses. CONF = confidence rating, O/U = percentage over or under confident.

the type of exposure, $F(1, 724) = 0.54$, $SEM = 2.39$, $\eta^2 = .00$. The two variables did not interact, $F(1, 724) = .332$, $SEM = 1.24$, $\eta^2 = .00$.

Confidence/accuracy relationship. Across conditions, participants were more confident in hits ($M = 59\%$) than misses ($M = 35\%$), $F(1, 724) = 79.66$, $SEM = 1.24$, $\eta^2 = .10$. O/U percentages are reported in Table 5. Overconfidence did not differ following neutral refreshing $F(1, 724) = 2.60$, $p > .10$, $SEM = 1.71$, $\eta^2 = .00$, nor depending on the type of product exposure, $F(1, 724) = 1.39$, $p < .238$, $SEM = 1.84$, $\eta^2 = .00$. Again, there was not an interaction between the two variables, $F(1, 724) = .961$, $p < .327$, $SEM = 1.71$, $\eta^2 = .00$.

Calibration curves were also constructed for neutral and control refreshing conditions (see Figure 7). These curves reveal nearly identical patterns of calibration, suggesting that neutral refreshing does little to alter the subjective experience of witnesses as they make product identifications.

Discussion

Refreshing with photos of all products appearing as options on the ultimate product identification test greatly reduced the effects of refreshing on responding. Experiment 3 concluded that the presence of correct photos in refreshing arrays overshadowed the effect of competing distractors, pushing accuracy to near ceiling levels following only ten minutes and significantly raising the rate of correct identification even after a week's time. The findings of Experiment 5 suggest that the presence of the correct photo in the refreshing array was not itself sufficient to significantly elevate correct identifications. This challenges the refreshing hypothesis suggesting the correct product is jogging memories while supporting the suggestibility hypothesis that both the positive and negative effects of refreshing noted in previous experiments resulted primarily from its suggestive nature.

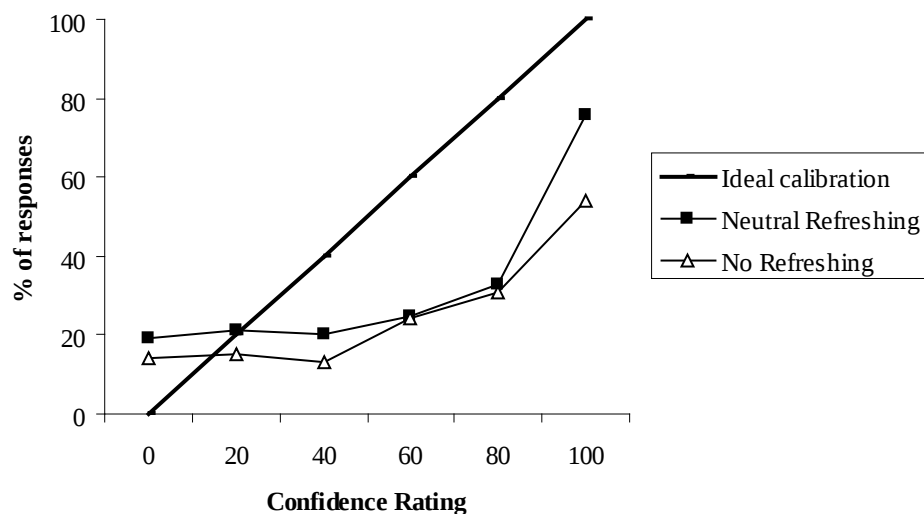


Figure 7. Confidence calibration curves for hits, Experiment 5

Though the manipulation of extended and close-up viewing was not of primary interest, photo refreshing had a slightly better chance of succeeding when participants

were provided with exposure opportunities that allowed them to focus on little-else besides the products themselves. Regardless of the film viewed, the benefits of viewing the correct product were not nearly so robust as those noted when only viewing two products.

The confidence calibration curves confirmed that viewing neutral refreshing arrays did little to alter the subjective experience of the product identification process. Hits were made with only slightly improved confidence following refreshing, while participants in both conditions remained largely underconfident in correct responses but overconfident in the experiment as a whole. Though it cannot be argued that neutral refreshing provides no help whatsoever to witnesses in product identification situations, the theory that refreshing achieves its effects primarily by suggestion, and not by jogging memories, is strongly supported.

Experiment 5 forces a modification of initial conclusions regarding the efficacy of photo refreshing. The ultimate goal of photo refreshing should be to increase the likelihood of a correct identification. Not only does it seem that the true product witnessed must appear in the refreshing array for this to happen, but that it must also appear when surrounded by a limited number of distractor photographs. This presents a problem for real-world identification scenarios, where the true product is unknown. Refreshing witnesses with enough products to reliably include the true product will likely involve drowning out any recollection of the true product with the sheer number of options. However, by refreshing with a limited number of photographs one not only risks excluding the true product from the array, but also risks biasing the witness towards one of the few products presented.

While many differences exist between real-world scenarios and these attempts at experimental replication, what makes post-event information suggestible to witnesses is likely the same. The most basic finding of this study is that product memories are poorly maintained—this is true whether a week has passed in an experiment or months and years have passed in the real world. Despite this, the entire purpose of an event (be it an experiment or a discovery session with an attorney) is to make product identifications, and witnesses realize that their failure to do so would make the experiment or the discovery session less beneficial. In these situations, simple demand characteristics likely dictate that the expected course of action is for witnesses to make the identifications using the information available at the time of retrieval.

Though the stated purpose of refreshing is not to distort or bias responding, both experimenters and attorneys are not surprised when this is the result (see Brickman, 2004). If photographs can act as powerful sources of suggestion to participants in a laboratory experiment, imagine the impact in a real-world case involving real loss and the possibility of real compensation. The proper identification in a liability trial could net a lucrative lawsuit worth a substantial sum of money—one necessary to offset prolific medical bills and in some sense atone for a life cut short by disease. The gravity of the two situations could not be more opposite, yet despite the obvious differences between the laboratory and the real world, the processes of suggestion and demand characteristics are similar. Consider again the Narcolite example discussed in the introduction, in which changes in the demand characteristics fundamentally changed the products identified.

CHAPTER THREE

General Discussion

The fragility of eyewitness memory and the resulting lack of reliability in witness testimony established by Loftus and others has gained widespread acceptance in both scholarly and legal arenas. As a result, the testimony of memory experts in criminal cases involving eyewitness identifications is now commonplace (Sporer et al., 1995). Additionally, extensive guidelines offering recommendations for police conducting lineups of suspects for witnesses in criminal cases have been prepared by the U.S. Department of Justice (1999). Testimony preceded by refreshing tactics failing to conform to these guidelines is often considered unreliable in criminal trials, in large part due to the increasing recognition of eyewitness errors leading to false convictions (see Loftus, 2004). In conducting the present experiments, I continued to explore the effects of post-event suggestions and the passage of time—two of the greatest threats to accurate eyewitness memories (Wells & Loftus, 2003)—but did so in a product identification paradigm.

There are several theoretical explanations of the findings of the studies. Classic ideas on misinformation and false memory creation can explain a majority of the findings, and in certain ways this project was inspired by those findings (see Loftus & Hoffman, 1988). In many ways, however, this study could be simply regarded as an exploration into how people make identifications when they are uncertain, especially when they are attempting to remember incidentally-encoded information. Both lines of

reasoning are discussed below, as well as the implications of this research on civil law. Finally, several limitations of this research are discussed—many of which make the findings all the more intriguing.

For the sake of this discussion, it is important to clearly define a few critical terms. Experiment 2 utilizes what is classically regarded as *misinformation*—post-event suggestions that contradict the original event witnessed. The errors caused by these inaccurate suggestions are *misinformation effects*, by definition. The photo refreshing experiments, in contrast, provided both correct and incorrect information, and did so in a non-authoritative manner. Thus, distortions caused by these suggestions are not true misinformation effects. Participants were free to accept or reject these suggestions, whereas the misinformation in Experiment 2 was stated as if factual. Both the classic misinformation effects (Experiment 2) and the effects of photo refreshing (Experiment 3-5) demonstrated effects of post-event suggestion. Thus, both types of experiment ultimately demonstrated suggestibility, regardless of whether the information presented was correct or not. Therefore, suggestibility is ultimately defined not by the content of a suggestion, but by its impact on subsequent responding.

There are two primary schools of thought on how misinformation achieves its distorting effects. Loftus et al. (1978) claimed that post-event suggestions distort responding by altering or eradicating the original memory traces. Others maintain that it is not the original memory traces that are affected by misinformation, but rather the processes of retrieval (McCloskey & Zaragoza, 1985; Morton et al., 1985). Chandler et al. (1991) argued that both the original trace and the post-event suggestion are recalled at the time of retrieval, but witnesses mistake the suggestion for the event and commit

source confusion errors. McCloskey and Zaragoza (1985) offered several “non-impairment” explanations of misinformation, including the possibility that both the event and the suggestion are remembered distinctly, but that the suggestion of the experimenter is simply better-trusted than the original memory trace. This idea will be important in interpreting some of the findings of Experiment 2. Loftus and Hoffman (1988) provide an overview of the classic misinformation studies and theories.

While these theories focusing on memory impairment and non-memory impairment have been the predominant focus of the field, Loftus and Hoffman (1988) acknowledged that a third way misinformation can affect responding is by filling the void left by the absence of an original memory trace. When individuals have no memory for the information requested, they may confidently adopt post-event suggestions as their memory of the event. Therefore, the theory that best describes many of the phenomena of this study is likely not eradication, impairment or source confusion, but rather memory adoption or memory acceptance (see McCloskey & Zaragoza, 1985).

One important distinction between the present study and the ones investigating the theories above is the inherent difficulty of the memory task. It is worth emphasizing that participants in the present experiments were inherently skeptical of their ability to recall product information. For most, there was an immediate lack of confidence in their ability to complete the task, even following the short ten minute delay. Experiment 1 confirmed that the ability to recognize brands from the film was at chance following the one week delay, and below 50% following a delay as brief as five minutes. This differentiates the present experiments from prior misinformation studies not only in the

durability and qualities of the memories being distorted, but also in the participants' *perception* of their abilities to remember.

Source attribution errors have been rejected as an explanation throughout this paper because it seemed unlikely that participants were confusing exposure to the film and exposure to post-event information—in most cases these events were overtly labeled as distinct. However, it does appear participants were often confused about what prompted their sense of familiarity with certain brands when ultimately identifying products. In this sense, source confusion was taking place. Especially in Experiment 2, the high rates of confidence in false alarms suggest that, even if participants did realize they were borrowing embedded suggestions from the questionnaire, they mistook the familiarity with these names at testing time as being sponsored by exposure to the product in the film, a form of source confusion described as hindsight bias, or the “knew-it-all-along” effect (see Arkes et al., 1988). Dunning and Stern (1994) suggest that familiarity causes certain answers to “pop-out” at participants when they scan the available options on a multiple choice questionnaire. This distinctiveness is then mistaken for evidence that the option was witnessed during the critical event. This appears to have occurred in the present study without post-event suggestions of any kind in Experiment 1, and occurred to a much greater extent when already familiar brands were suggested following the event in Experiment 2. Therefore, while watching the film and reading the questionnaire were likely recalled as distinct events at the time of testing, the senses of familiarity created by these separate occasions appears to be melded together and attributed to the event rather than what followed. In this sense, overconfidence was likely driven by source confusion.

Photo refreshing differs from classic misinformation exposure in that it not only introduces post-event information overtly, but does so in a non-authoritative sense. Participants realized they were receiving at least one incorrect suggestion in each refreshing slide in Experiments 3 and 4, and therefore were granted a clearer opportunity to judge the information with skepticism. While this combination of factors did not reduce the rate at which post-event suggestions were accepted following two-item refreshing, it did reduce the confidence with which errors were made, while concurrently increasing the confidence with which correct identifications were made. Correct responses were made with more certainty than misses in all photo refreshing conditions, suggesting that less source confusion regarding familiarity took place.

Following Experiment 3, two conclusions regarding the role of two-product photo refreshing seemed viable. On the one hand, presentation of the correct product did appear to jog memories of the original event—accuracy was boosted to near ceiling in ten minute delay experiments. On the other hand, presentation of incorrect products had an equally powerful effect on responding—one of the two incorrect products was identified 74% of the time following a ten minute delay, and a stunning 86% of the time following a one week delay. Was the presence of the true product in correct photo conditions sufficient to actually jog memories that had otherwise slipped beyond the participants' ability to recall? Or was the previously-witnessed brand identified simply because its suggestion made it more plausible than other options, while only a vague, subconscious sense of familiarity concurrently caused it to “pop-out” more than the brand name of the product suggested alongside it?

Experiment 5 strongly suggests that the primary utility of photo refreshing following a one week delay is to provide suggestions that can be adopted when original memories of the event have not been maintained. The benefits of non-suggestive refreshing were sufficiently minimal to make clear that the mere presence of the correct product in the refreshing array was rarely jogging existing memories. Not only did adding options to the refreshing array in Experiment 5 strongly diminish the benefits of viewing the correct product, it also eliminated all distorting effects of post-event suggestions, making the pattern of responding similar to that noted in Experiment 1³.

Experiments 3 and 4 demonstrated that refreshing participants with only a selection of possible answers strongly biases them towards identifying one of the refreshed products. While the correct product was preferred in these situations, refreshing with exclusively incorrect products did not reduce the rate at which one of the refreshed products was selected. This is inconsistent with the assumption that the sole purpose of photo refreshing is to jog existing memories—if this were the case, viewing two incorrect photos should not impair accuracy, and should in fact bolster the rate of identifying the non-refreshed products. Indeed, viewing two incorrect products should help eliminate them from consideration when making product identifications; in contrast, viewing two photos seemed to make the suggested brands the exclusive focus of consideration during the identification process, regardless of their accuracy.

Though it does not require much elaboration, one of the most important points of this research is that product identification was incredibly difficult—Experiment 1

³Familiar false alarms were not reported in Experiment 5 because many of the condiments in the extended exposure film did not have a brand clearly identified as “familiar.” For several of the products, two or three brands were equally identified during normative questioning. Despite this, the pattern of misses noted in Experiment 5 was comparable to that seen in Experiment 1.

established this in its simplest terms. This was true for the most part because product information was likely never encoded to begin with. Look no further than Nickerson's and Adams' classic study (1979) examining the inability of most Americans to correctly recognize the layout of the penny—a coin most people handle at least once a day. This failure to recognize the features of the penny does not result from our having forgotten the layout—it was simply never learned in the first place, because encoding those features was not necessary. In most cases, product information was likewise never learned to begin with, because remembering a brand name is not essential to using the product. More than exploring eyewitness memory, these findings demonstrate how participants make difficult memory decisions in the absence of reliable memories on which to base them.

The “non-existent photo” condition of Experiment 4 produced strong evidence supporting the theory that photo refreshing in a product identification situation is largely suggestion. Chiefly, the fabricated different brand/different container products (including Smuckers barbecue sauce, Hormel relish, Sara Lee mayonnaise, and Hillshire Farms ketchup) were identified a remarkable 26% of the time following eight seconds of photo refreshing. These contrived products were previously included as options on questionnaires for control groups in a pilot study that were exposed to no refreshing. In total, participants made over six hundred product identifications in the pilot study—not one was a non-existent product. These errors are not made as a result of blind guessing or a simple lack of attention, but solely as the result of suggestive photo refreshing.

The original purpose of Experiment 4 was to investigate the role of perceptually-driven processes in product identification. The different brand/different container

condition was constructed in the interest of fully-balancing the experimental design, and was included mainly as a control group with the expectation that participants surely would not identify non-existent products comprised of unseen brands and unseen containers. The rate of identifying one of the two photos in this condition (a stunning 83%), suggests that product identification following poor encoding or insufficient rehearsal results from a combination of suggestion and demand characteristics. The mere fact that the experimenter has selected the products for the refreshing array appears to provide participants with their best reason for making an identification.

The confidence data supports one steadfast conclusion: Photo refreshing allowed participants to distinguish between true memories and those based on suggestion—classic misinformation did not. Participants in Experiment 2 confused their familiarity with the suggested brands as evidence that they had seen them in the film, and were more confident in false identifications than true ones. However, participants in photo refreshing experiments attributed higher confidence to their hits, even when two incorrect photos were used in refreshing. When participants had little or no memories of product exposure (as seems to have been the case quite often), they appear to have utilized the post-event information in Experiments 3 and 4 to narrow the possibilities, and then selected one of the brands displayed in the photographs. This produced different patterns of confidence assessment and a reduced reliance on pre-existing familiarity. Curiously, unfamiliar false alarms were more common than familiar false alarms following familiar/unfamiliar refreshing in the ten minute delay experiment. More predictably, the opposite was true in the one week delay experiment. This discrepancy is difficult to explain, but it supports the idea that when the 50/50 mentality of responding is adopted,

other sources of information (such as pre-existing familiarity, and even original memory traces) are given less credence. Ultimately, it can be assumed that the few hits made confidently following refreshing with two incorrect products (23% following ten minutes, 6% following one week) reflect the rate at which actual memories or lucky guessing can sponsor accurate responding following exclusively inaccurate post-event suggestions. This reveals the inherent problem with product identification—very few memories are encoded and rehearsed sufficiently to survive the delay between exposure and subsequent relevance, much less suggestive refreshing with unseen brands

Though this study does not provide ample evidence to support it, the conclusion that many of the product identifications in this study were made in the absence of any meaningful memories of the products originally viewed would not be far-fetched. Lacking any tangible brand memories, participants were forced to rely entirely on other sources of information, such as familiarity, suggestion or strategic guessing. In this scenario, pre-existing familiarity and post-event suggestions would be utilized exclusively to formulate plausible guesses. In this sense, participants would not adopt post-event suggestions as their memories, but rather would provide an informed guess on the questionnaire and leave the experiment session still convinced they have no memories of what they saw in the film.

A future experiment that could easily test this theory would ask participants to fill out the product identification questionnaire a second time following a considerable interval. If they failed to replicate their answers in the absence of a second round of refreshing, it would suggest that the misinformation has not actually been adopted as a memory. If identifications were repeated, it would indicate that photo refreshing does in

fact create lasting memories that either replace pre-existing traces or fill gaps in recollection. A greater tendency to commit familiar false alarms would indicate that participants were again relying on pre-existing familiarity as their basis for responding. Other experimental manipulations which would flesh out this line of refreshing involve the utilization of a non-forced choice test for product identification. Including a “don’t know” option on the questionnaire would give participants an opportunity to admit that they do not remember the product in the film, even following refreshing. If the presence of the “don’t know” option did not alter the findings of the present study, it would support the theory of memory creation. If participants readily conceded that they did not know, it would support the hypothesis of suggestibility, indicating that refreshed products will only be selected when an identification is required. A third experiment that would fully test the hypothesis of suggestion would refresh and test participants a second time following the initial identification session. At the second session, participants would be refreshed with a separate subset of products. If identification differed at test two compared to test one, it would suggest that the previous refreshing did not create lasting memories, but only provided suggestions applicable to the present identification. However, if participants resisted the influence of the second round of refreshing and replicated their initial identifications, it would support the hypothesis of memory jogging or memory creation.

Implications for Civil Law

While the lengthy debate above regarding the subjective meaning of confidence ratings and the theoretical mechanisms driving refreshed identifications in these situations may intrigue cognitive psychologists, the meaning is unambiguous when it

comes to the courtroom. Witness confidence is a key factor assessed by jurors when attempting to determine the credibility of a witness (Koegh & Markham, 1998, Wells, 1993). Across the board in this study, if participants were exposed to biasing post-event information, they were more likely to commit confident false alarms than respond correctly. Similar effects are likely to occur with witnesses in civil liability situations following persuasive refreshing. This is particularly likely if the brand names of the products involved were of no specific interest to the witnesses.

Contrary to the refreshing hypothesis, these results suggest that witnesses use photo refreshing to create rather than jog memories. If only a subset of products was used in refreshing, one of these products was likely to be selected with high confidence, regardless of whether this product was present at the time of critical exposure. Finally, in the absence of clear memories, witnesses will likely select the product based on indirect processes, such as name familiarity or plausibility.

Limitations

No paper can identify all possible factors involved in product identification. One important distinction between actual civil cases and brief psychology experiments is the potential benefit to witnesses for their identifications. In real life, witnesses are likely very ill—in many cases lethally so. This creates a profound motivation to produce an identification that cannot only be acted upon legally, but also generate sufficient monetary compensation to atone fully for pain and suffering, and guarantee that surviving family is provided for adequately. Although it is possible to study and control demand characteristics in the laboratory, one cannot experimentally mimic such powerful motivators.

While this distinction in many ways limits the present research, in others it makes the findings all the more startling. Even in situations with relatively little motivation to answer in one fashion or another (aside from perhaps an innate desire to perform well on a test), participants still readily succumbed to experimental suggestions made both overtly and covertly. This suggests that the process of incorporating suggestions into response patterns does not result exclusively from the desire to enhance financial compensation, nor from a lack of interest by laboratory participants. Rather, the incorporation of suggested information is likely part of a decision-making process that emerges whenever memory is needed for events that have not been well encoded or sufficiently rehearsed. This suggests that a tendency to incorporate unlikely suggestions into memories, whether during an experiment or a discovery session, is likely not motivated by desperation or even simple greed—it appears to be a cognitive process that mediates responding regardless of the perceived result of the identification.

Another obvious challenge in this type of research is experimentally replicating multiple product exposures that in the real world may take place incidentally over weeks, months, or even years. Achieving such long-term exposure within the confines of an experiment is nearly impossible. Further, research suggests that witnessing an event live sponsors stronger memories than does watching a video tape of the same event, a trend analogous to actor/observer bias (Roebbers et al., 2004). However, other research conducted in our laboratory (Colby & Weaver, 2006) found that observers had better brand memories than actors. The fact remains that there is no experimental method for truly incidental product exposure—participants are either forced to interact directly with the product in question, or asked to watch videos containing deliberately placed shots of

the products. Real world exposures provide numerous opportunities for brand name exposure, but rarely force or impel witnesses to take notice of brands the way close-up shots might in the product exposure film.

In an attempt to expose participants to products in a more natural manner over repeated instances, I placed a beverage can at the front of a classroom where a lab class met once a week. Twenty-one students came to class twelve times in this room during the course of the semester, with each class lasting between 30 and 45 minutes. These students were therefore incidentally exposed to the can for six to seven hours over the course of four months, with no mention of the can ever made during class. On the last day of class, the can was removed and students were offered extra credit on their final quiz if they could free recall the brand of the can that had sat on the chalkboard ledge all semester. While most students acknowledged remembering that a can had been present, only 4 of 21 remembered that it was a bright blue can of Deja Blue water. I chose this brand because it was fairly unusual and would almost certainly not be guessed by anyone. Most students (10) claimed it was a Dr. Pepper can, as this is the soda sold exclusively by Baylor and represented for the setting the most popular, and therefore most familiar, brand of soda. These findings closely resemble what was noted in Experiment 1, a 20% hit rate and a familiar false alarm rate approaching 50%. Though these are merely anecdotal data, they are consistent with the assumption that exposing participants to products via a short film is analogous to more incidental exposure opportunities spaced over weeks and months.

Finally, a mock trial study exploring the impact of sharing the findings of this study with jurors was attempted, but thus far has not gotten off the ground. The reasons

for this failure reveal further limitations of the capability of this research to generalize to real world situations. In this study, participants read about a girl who had become severely ill after spilling a bag of flour containing a mysterious toxin. Though the flour bag was thrown away and she could not initially remember the brand of flour, she was able to make a confident identification after meeting with her attorneys, who used photobooks to refresh her memory. After reading this initial synopsis, 53% of participants indicated that if the trial ended at that point, they would certainly side with the flour company over the plaintiff. Another 48% leaned towards siding with the company at this point, while only 9% leaned toward siding with the plaintiff following a reading of the initial synopsis. None of the 115 participants initially indicated they would certainly side with the plaintiff following the initial reading, while 91% favored absolving the company of any responsibility. This left few minds to be changed by the testimony of an expert witness, and essentially left nothing to measure. In an attempt to move the baseline reaction, I changed the story so that the victim had not just gotten ill and subsequently recovered, but was actually afflicted with a lethal illness and had less than three months to live. This had little effect—still close to 80% of participants sided with the defendant company following the initial synopsis.

This surprising finding reveals another way in which experimental replication of product liability trials is difficult. The results suggest that when the physical presence of pain, suffering, and loss is removed from the situation, jurors are better able to judge a case solely on the facts presented. Participants knew this was not a real case, and that their judgment would not actually send a dying girl home empty-handed. The overwhelming refusal to award the plaintiff damages in the mock trial suggests that real-

world juries may overlook certain facts and subdue certain skepticisms in order to do what appears to be the right thing, a phenomenon that has been described as distributive justice (see Sommer, Horowitz, and Bourgeois, 2001). Jurors in most asbestos lawsuits are faced with a difficult decision. Many of the plaintiffs in these cases have clearly been exposed to asbestos, and most are slowly dying painful deaths as a result. Jurors also likely realize that the company on trial almost certainly did not manufacture the product that caused the illness—the findings of the mock trial study seem to support this. The question ultimately answered by jurors is likely not the proper one: “Is this company liable for the plaintiff’s illness?” Rather, jurors likely ask themselves: “Does the plaintiff deserve compensation for their illness and suffering?” The good of rewarding a dying cancer victim who clearly deserves restitution seems to often outweigh the injustice of penalizing a likely innocent company. Jurors often justify these decisions by arguing that corporations have plenty of money to spend, while individual families can be destroyed by excessive medical bills—a phenomenon described by the wealth redistribution hypothesis (see Turner et al., 1994). In a mock trial, with no real pain or suffering to consider, participants are seemingly able to consider the proper question of company liability without the burden of potentially sending a legitimate victim home with no compensation.

Final Conclusions

Even though these experiments are not perfect replications of real-world product identification scenarios, they paint an unambiguous picture about the inherent difficulty of product identification and the processes used in scenarios requiring poorly encoded memories for incidental information. Following a week long delay, witnesses were only

able to recognize previously viewed product brand names at chance rates. In the absence of well-encoded memories, participants confidently adopted memories sponsored by pre-existing familiarity, embedded written suggestions and suggestive photo refreshing. Confidence only proved a reliable predictor of accuracy when participants received no post-event suggestions whatsoever. Finally, refreshing designed to be non-biasing failed to significantly jog memories. In most cases, product memories likely could not be jogged because they were never sufficiently encoded in the first place. Attempts to do so were likely to induce confidently-held false memories rather than true recollections.

APPENDICES

APPENDIX A

Style Questionnaire, Film 1

1) How often do you bake?

Frequently

Once a week

Once a month

Rarely

Never

2) What was your overall impression of Anjie Smith?

Loved her

Hated her

She's strange

Indifferent

3) Did you feel that her attitude was condescending?

Yes

No

4) Would you have enjoyed the program more if it had been filmed in a kitchen?

Yes, it was stupid as it was

No, it was fine as it was

Nothing you could do would make me enjoy that program

5) How would you compare your reaction to that of the students in the film?

Similar

More enthused

I would've left sooner

6) What would've made the film better?

7) What parts of the film did you like? Would you attempt to use Anjie's recipe?

8) If you wanted to bake cookies, would watching this film be sufficient to understand how to do it? Why or why not?

APPENDIX B

5-AFC Test, Film 1

Last four digits of your ID number _____

Please answer the following questions about the film you just watched. When asked, also indicate the confidence you have in the accuracy of your answer by circling one of the percentages. If you have no idea and are completely guessing, please circle 0%. If you are absolutely certain, please circle 100%. Use the other percentages to indicate confidence between these two extremes.

1) What brand of **sugar** was used in the recipe?

Holly H-E-B Imperial Domino Great Value

How confident are you? 0% 20% 40% 60% 80% 100%

2) What brand of **flour** was used in the recipe?

Gold Medal King Arthur Hill Country Fare Martha White W-R

How confident are you? 0% 20% 40% 60% 80% 100%

3) What brand of **baking soda** was used in the recipe?

Great Value Hill Country Fare Arm & Hammer Albertson's Best Yet

How confident are you? 0% 20% 40% 60% 80% 100%

4) What brand of **salt** was used in the recipe?

Alessi Morton Best Yet La Fina H-E-B

How confident are you? 0% 20% 40% 60% 80% 100%

5) What brand of **baking powder** was used in the recipe?

Hill Country Fare Great Value Best Yet Clabber Girl Calumet

How confident are you? 0% 20% 40% 60% 80% 100%

6) What was the chef's name? _____

7) What brand of chocolate chips were used? _____

8) Did the chef wear glasses? Yes No

APPENDIX C

5-AFC test, Film 1

Last 4 digits of ID number _____ Exp 8c-1a **READ THE INSTRUCTIONS!**

Please answer the following questions about the film you watched earlier. **Please circle a response even if you are uncertain.** Circle one of the percentages to indicate your confidence in your answer. If you are completely guessing, circle 0%. If you are absolutely certain, circle 100%. Use the other percentages to reflect confidence between these two extremes. **It is very important that you answer each question.**

1) What brand of mayonnaise did the artist use?

	Kraft	Hellmann's	Sara Lee	Great Value	H-E-B	
How confident are you?	0%	20%	40%	60%	80%	100%

2) What brand of mustard did the artist use?

	Gulden's	French's	Nestle	Plochman's	Colman's	
How confident are you?	0%	20%	40%	60%	80%	100%

3) What brand of barbecue sauce did the artist use?

	Bullseye	Kraft	KC Masterpiece	Tyson	Sweet Baby Ray's	
How confident are you?	0%	20%	40%	60%	80%	100%

4) What brand of peanut butter did the artist use?

	Pillsbury	Jif	Reese's	Skippy	Peter Pan	
How confident are you?	0%	20%	40%	60%	80%	100%

5) What brand of relish did the artist use?

	Best Maid	Hormel	Mt. Olive	Vlasic	Heinz	
How confident are you?	0%	20%	40%	60%	80%	100%

6) What brand of ketchup did the artist use?

	Hillshire Farms	Del Monte	Heinz	Hunts	Great Value	
How confident are you?	0%	20%	40%	60%	80%	100%

7) List as many of the artist names as you can remember:

_____	_____					
_____	_____					
How confident are you?	0%	20%	40%	60%	80%	100%

APPENDIX D

Style survey, Film 2

Last 4 digits of ID number _____ Exp 5 **READ THE INSTRUCTIONS!**

Answer the following questions about the film you just watched truthfully and in as much detail as possible.

1. Your gender: M / F 2. Your age _____ 3. Your classification: Fresh / Soph / Jr / Sr

Indicate your agreement with the following statements using the 1 – 9 scale.

I enjoyed the news report.	1	2	3	4	5	6	7	8	9	
strongly disagree		disagree				no opinion		agree		strongly
agree										
I enjoyed Anjie Smith's style of reporting.	1	2	3	4	5	6	7	8	9	
strongly disagree		disagree				no opinion		agree		strongly
agree										
I might use food for art projects in the future.	1	2	3	4	5	6	7	8	9	
strongly disagree		disagree				no opinion		agree		strongly
agree										
Using food for art is wasteful and irresponsible.	1	2	3	4	5	6	7	8	9	
strongly disagree		disagree				no opinion		agree		strongly
agree										
I thought the program was entertaining.	1	2	3	4	5	6	7	8	9	
strongly disagree		disagree				no opinion		agree		strongly
agree										
I was moved by the film's depiction of evangelism.	1	2	3	4	5	6	7	8	9	
strongly disagree		disagree				no opinion		agree		strongly
agree										

Answer the following questions as completely as possible.

How long did the film last? _____

Which bible verse was cited as an inspiration for food art? _____

What did Aaron Patterson like to paint? _____

Who owned "Condiment Creations?" _____

How often do you purchase condiments at the store? Never Occasionally Regularly Often

Would you consider buying food art if you encountered a vendor like the one on the Bear Trail?

REFERENCES

- Arkes, H. R., Faust, D., & Guilmette, T. J. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73*, 305-307.
- Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*, 1-29.
- Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Quarterly Journal of Experimental Psychology, 16*, 318-324.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55-68.
- Biederman, C., Korosec, T., Lyons, J., & Williams, W. (1998, August 13). Toxic justice. *The Dallas Observer*.
- Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology, 63*, 545-552.
- Brainerd, C. J., & Reyna, V. F. (1990). Gist is the gist: The fuzzy-trace theory and new intuitionism. *Developmental Review, 10*, 3-47.
- Brickman, L. (2004). On the theory class's theories of asbestos litigation: The disconnect between scholarship and reality. *Pepperdine Law Review, 31*, 33-170.
- Carroll, S. J., Hensler, D. R., Gross, J., Sloss, E. M., Schonlau, M., Abrahamse, A., et al. (2005). Asbestos Litigation. Santa Monica, CA: Rand Corporation.
- Chandler, C. C., Gargano, G. J., & Holt, B. C. (2001). Witnessing postevents does not change memory traces, but can affect retrieval. *Applied Cognitive Psychology, 15*, 3-22.
- Colby, M. A. E., & Weaver, C. A., III. (2006). Comparing eyewitness memory and confidence for actors and observers in product identification situations: Extending findings and methodology from criminal justice. *Applied Psychology in Criminal Justice, 2*, 145-162.

- Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 142-150.
- Dunning, D. & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitnesses via inquiries about decision processes. *Journal of Personality and Social Psychology*, 67, 818-835.
- Ebbinghaus, H. (1913/1885). *Memory: A Contribution to Experimental Psychology* (H. A. Roger & C. E. Bussenius, Trans.). New York: Teachers College.
- Frost, P. J., & Weaver III, C. A. (1997). Overcoming misinformation effects in eyewitness memory: Effects of encoding time and event cues. *Memory*, 4, 1-12.
- Garry, M., & Gerrie, M. P. (2005). When photographs create false memories. *Current Directions in Psychological Science*, 14, 321-325.
- Garry, M., & Wade, K. A. (2005). Actually, a picture is worth less than 45 words: Narratives produce more false memories than photographs do. *Psychonomic Bulletin & Review*, 12, 359-366.
- Godden, D., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology*, 66, 325-331.
- Gordon, P. C., & Holyoak, K. J. (1983). Implicit learning and generalization of the mere exposure effect. *Journal of Personality and Social Psychology*, 45, 492-500.
- Hampel, P. (2005, May 10). Defense attorneys grill witness in asbestos trial. *The St. Louis Post-Dispatch*.
- Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-340.
- Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, 122, 139-154.
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32, 1-24.
- Koehn, L., & Markham, R. (1998). Judgments of other people's memory reports: Differences in reports as a function of imagery vividness. *Applied Cognitive Psychology*, 12, 159-171.

- Koriat, A. (1995). Dissociating knowing and feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124, 311-333.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349-370.
- Krug, K. S. (2004). Eyewitness memory in civil cases: Can witnesses correctly identify products used in prior situations? Unpublished doctoral dissertation, Baylor University, Waco, TX.
- Krug, K. S., & Weaver, C. A., III. (2005). Eyewitness memory and metamemory in product identification: Evidence for familiarity biases. *The Journal of General Psychology*, 132, 429-445.
- Lindsay, D. S. (1993). Eyewitness suggestibility. *Current Directions in Psychological Science*, 2, 86-89.
- Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 1077-1083.
- Lindsay, R. C. L., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness identification accuracy within and across situations? *Journal of Applied Psychology*, 66, 79-89.
- Loftus, E. F. (2004). Memories of things unseen. *Current Directions in Psychological Science*, 13, 145-147.
- Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory and Cognition*, 17, 607-616.
- Loftus, E. F. & Hoffman, H.G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, 118, 100-104.
- Loftus, E. F., Miller, D. G., & Burns H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19-31.
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, 25, 720-725.

- Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytical basis for JOLs? *Memory and Cognition*, 29, 222-233.
- McCloskey, M., & Zaragoza, M. S. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114, 1-16.
- Morton, J., Hammersley, R. H. & Bekerian, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, 20, 1-23.
- Nickerson, R.S. & Adams, J.J. (1979). Long-term memory for a common object. *Cognitive Psychology*, 11, 287-307.
- Penrod, S. D., & Cutler, B. L. (1995). Witness confidence and witness accuracy: Assessing the forensic relation [Special theme: Witness memory and law]. *Psychology, Public Policy, and Law*, 1, 817-845.
- Robinson v. A.W. Chesterton, et. al., No. 393414 (Court of Common Pleas, Cuyahoga County, Ohio 2001).
- Roebers, C. M., Gelhaar, T., & Schneider, W. (2004). "It's magic!" The effects of presentation modality on children's event memory, suggestibility, and confidence judgments. *Journal of Experimental Child Psychology*, 87, 320-335.
- Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist*, 45, 1043-1056.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 803-814.
- Schacter, D (1996). Searching for memory: The brain, the mind, and the past. New York: Basic Books.
- Sommer, K. L., Horowitz, I. A., & Bourgeois, M. J. (2001). When juries fail to comply with the law: Biased evidence processing in individual and group decision making. *Personality and Social Psychology Bulletin*, 27, 309-320.
- Sporer, S. L., Penrod, S., Read, D. & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315-327.
- Suarez, E. C. (2003). A princess in god's eyes: Cinematherapy as an adjunctive tool. *Journal of Psychology and Christianity*, 22, 259-261.

- Terrance, C. A., Matheson, K., Allard, C., & Schnarr, J. A. (2000). The role of expectation and memory-retrieval techniques in the construction of beliefs about past events. *Applied Cognitive Psychology*, 14, 361-377.
- Terrell, J. T., & Weaver, C. A., III. (in press). Eyewitness testimony in civil litigation: Retention, suggestion, and misinformation in product identification. *North American Journal of Psychology*.
- Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, 27, 1250-1260.
- U. S. Department of Justice (1999). Eyewitness evidence: A guide for law enforcement (No. NCJ Publication 178240). Washington, DC: U.S. Department of Justice/Office of Justice Programs.
- Wade, K. A., Garry, M., Read, J. D., & Lindsay, S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin and Review*, 9, 597-603.
- Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, 48, 553-571.
- Wells, G. L. & Loftus, E. F. (2003). Eyewitness memory for people and events. In A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology*, Vol. 11 (pp. 149-160). New York, NY, John Wiley & Sons, Incorporated.
- Winningham, R. G., & Weaver III, C. A. (2000). The effects of pressure to report more details on memory of an eyewitness event. *European Journal of Cognitive Psychology*, 12, 271-282.
- Zachary, R. A. (1986). Shipley Institute of Living Scale: Revised manual. Los Angeles: Western Psychological Services.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monographs*, 9, (2, Pt. 2).
- Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 934-945.