COMPARING EYEWITNESS MEMORY AND CONFIDENCE FOR ACTORS AND OBSERVERS IN PRODUCT IDENTIFICATION SITUATIONS: EXTENDING FINDINGS AND METHODOLOGY FROM CRIMINAL JUSTICE

M. Amanda Earl Colby
Charles A. Weaver, III

Baylor University

Numerous studies have demonstrated problems with the reliability of eyewitnesses in criminal justice setting. Many of the same concerns apply to eyewitnesses in civil cases involving product identification testimony, as in asbestos litigation. We examined differences in product identification memory between witnesses with differing levels of involvement, the effects of delay before testing, experience, pre-existing familiarity, and participants' self-reported confidence. Participants either mixed recipes (actors) or observed this mixing (observers) and later were tested about the brands used. Contrary to expectations, observers were slightly more accurate than actors, though all witnesses were influenced by pre-existing familiarity. Confidence was unrelated to accuracy in all conditions. Participants with more baking experience were more confident, but not more accurate, suggesting experience inflates confidence without improving accuracy. We discuss implications for matters of product identification testimony.

Correspondence concerning this article should be addressed to Charles A. Weaver, Baylor University, Box 97334 Department of Psychology and Neuroscience, Waco, TX 76798, Email: Charles_Weaver@Baylor.edu

The testimony of eyewitnesses often plays a vital role in determining the innocence or guilt of a defendant. Often the recollection of a single individual represents the crux of the prosecution's case, and thus considerable research has been conducted to help evaluate the reliability of witnesses' reports. With the growing scientific understanding that memory is continuously altered and reconstructed, the veracity of courtroom testimony is being reevaluated. Furthermore, expert testimony regarding the reliability of eyewitness testimony in criminal cases is relatively common. Ten years ago, Loftus (1996) reported that more than 100 memory researchers have testified in cases involving eyewitness memory. The number has certainly grown since then. Furthermore, current research indicates that much of the information presented by these experts in not within the common knowledge of jurors (see Kassin, Tubb, Hosch, & Memon, 2001; Schmechel, O'Toole, Easterly, & Loftus, 2006).

An escalating number of legal cases and lawsuits involve memories from the remote past. Controversy has raged over the reliability of recovered memories (ones that were once repressed) and whether they should be allowed as testimony in cases of alleged childhood sexual abuse (Terrance, Terrance, Matheson, Allard, & Schnarr, 2000). While there is frequently no way to prove conclusively whether recovered memories are accurate, evidence does exist demonstrating that memories of the distant past can be altered or created by suggestions, even if participants initially report no recollection of the event. Several recent articles illustrate the powerful effects exerted by retrieval cues (Lindsay, Hagen, Read, Wade, & Garry, 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. Finally, 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.

The overwhelming majority of studies in eyewitness memory involve criminal justice settings. Many of the findings are sufficiently understood and accepted that the Department of Justice has relied on this research to produce a guide for law enforcement regarding the treatment of eyewitness evidence (U. S. Department of Justice, 1999). Many of these same principles can be applied to cases involving civil disputes, particularly those in product liability lawsuits.

The question of the reliability of long-term memory for seemingly irrelevant events is growing increasingly pertinent in the courtroom because of the steady growth in the number of lawsuits hinging on claims of injustices some thirty or more years past. In particular, cases involving exposure to products containing asbestos, often brought by those suffering from asbestosis or mesothelioma, rely heavily on the victim's memory of being exposed to a specific asbestos-containing product. Little research has been conducted on very long-term memory for insignificant details such as the brand names of products people might have used years before (see Krug & Weaver, 2005; Terrell & Weaver, 2005; 2006, for exceptions). In most cases, written records that could confirm the presence or absence of asbestos-containing materials are unavailable, and thus the witnesses' recollections of working in its presence become the essence of the case (Weaver, 2004; Weaver, Terrell, & Holmes, 2006). Many of those victims sued companies they believed responsible for their illnesses, often winning large settlements.

A recent study by the Rand Corporation estimates through the end of 2002, nearly three-quarters of a million claims were filed for asbestos-related injuries, with damages totaling about $70 billion (Carroll et al., 2005). As a result, more than 70 companies have filed for bankruptcy, including many of the largest asbestos manufacturers such as Johns-Manville; between 1986 and 1998 Johns-Manville was estimated to have paid more than $1.8 billion in settlement claims and verdicts (Biederman, Korosec, Lyons, & Williams, 1998). Since the bankruptcy of these large firms, smaller asbestos manufacturers or companies that used small amounts of asbestos in their products have become the targets. Furthermore, those initiating lawsuits are less likely to be those who used asbestos directly, such as insulators, and are more likely to be "house-builders and roofers, supervisors, even office workers and supply clerks" (Biederman et al., 1998, ¶ 76)(Biederman et al., 1998). Much of the testimony offered in product liability cases involves passive activity—that is, many individuals are bringing suit against manufacturers of products *they saw others use*, as opposed to products they used themselves. This is similar to the actor-observer distinction offered in traditional criminal eyewitness memory research. Loosely defined, the actor-observer distinction predicts that actors (those actively involved in an event or the one speaking during an interaction) will attend to or focus on different aspects of an event or interaction than will observers. The majority of the research done on this phenomenon, however, is focused on interpersonal interactions.

Research on actor-observer distinction in eyewitness situations has been conflicting. For example, Hosch and Cooper (1982, as cited by Yuille, Davies, Gibling, & Marxsen, 1994) found that in criminal cases the victims (actors) were more accurate in identifying the criminal than were uninvolved witnesses (observers). However, Kassin (1984) found the opposite. In their own research at the police training school, Yuille and colleagues found that just one week following the staged event actors and observers displayed no difference in the amount of information recalled. However, after 12 weeks active participants remembered more.

Ihlebaek, Love, Eilertsen, and Magnussen (2003) had participants either participate in a staged robbery or watch one of these staged robberies on video. Participants who had viewed the video performed better on all aspects of their 12 question evaluation. Davies and Alonso-Quecuty (1997) found similar results. Participants viewed videos of cultural events with instruction to half of them to view it from the perspective of an observer whereas the other half were told to adopt the perspective of an active participant. Those in the observer role had more accuracy and fewer errors than participants, which was consistent with their hypothesis: Davies and Alonso-Quecuty expected observers to focus on the event outside themselves and therefore encode external (or objective) details whereas the participant perspective lends itself to more internal, or subjective, memories. Thus, because most questions about an event are aimed at determining what actually transpired rather than how someone felt about it, observers tend to do better because they encode more of this sort of information.

Interestingly, almost all evidence concluding observers' memory is superior is from experiments in which videos are used as the source of observation rather than live events. This has led many researchers to believe that studies conducted in laboratories on eyewitness memory inflate the accuracy of memory and that more ecologically valid research needs to be conducted (Behrman & Davey, 2001; Ihlebaek et al., 2003; Woolnough & MacLeod, 2001). One conclusion that can be drawn from the research on level of involvement is that witnesses close to the scene both remember more details and have greater accuracy than witnesses more distanced (Ihlebaek et al., 2003).

Finally, an important area is eyewitness memory research involves the role of familiarity in addition to (or sometimes instead of) recollection (Searcy, Bartlett, & Memon, 1999; Sherman, Groom, Ehrenberg, & Klauer, 2003). The classic finding of "unconscious transference" in eyewitness identification demonstrates possible misleading effects of familiarity (see Ross, Ceci, Dunning, & Toglia, 1994). Pre-existing familiarity with people or things to be identified can lead to errors, illustrating what Johnson and colleagues have called "source monitoring" problems (Johnson, Hashtroudi, & Lindsay, 1993).

The present research extends findings from eyewitness studies in the criminal justice context. It was designed to investigate these differences in situations more closely approximating those in product liability cases. Rather than having witnesses view a simulated crime, we had witnesses use different kinds of baking products in a simulated kitchen. Just as witnesses in product liability are asked to recall specific brand names, we tested witnesses' memories for the brands of products they used while "baking." Specifically, we investigated the effects of different involvement levels (participants with direct involvement being actors and indirect involvement being observers) as well as different retention intervals (five minutes or one week) on correct responses, false alarms, and subjective confidence. In addition, baking experience was examined as a covariate.

METHOD

Krug and Weaver (2005) developed a procedure to investigate memory for product identification. They used recipe mixing of common household ingredients to expose participants to a variety of products and brands, without informing them that they would be asked to remember the ingredient brands. Participants are later asked to identify what products they used and rate their confidence in their decision.

Participants

Participants were recruited from Baylor University's participant pool and offered extra credit for participating. A total of 100 college-age participants, both male and female, were recruited and asked to participate in a two-session study, returning for the second part exactly one week following the first session.

Neither part of the study took more than 30 minutes. Participants were assigned to conditions randomly, subject to the constraints described below, ensuring that approximately equal numbers of males and females were assigned to each group. Participants were tested individually or in pairs. Following both APA and Baylor University Institutional Review Board guidelines, the (minimal) risks of participation were explained, participants were informed of their rights as participants, and were told that individual data would be collected in such a manner as to preserve anonymity. Also following APA and University guidelines, alternative forms of credit were available to those who declined to participate. In addition to verbal descriptions, participants read a form explaining their rights, and any questions were addressed. Finally, participants were asked to sign the printed form explaining their rights, and indicating their consent to participate.

Procedure

As this experiment was an exploration of differences between actors and observers, participants were usually tested in pairs. Because this was not always possible, occasionally participants were tested separately. In the case of pair testing, the first participant to arrive was asked to pick a number between one and ten. If a number between one and five was picked, the participant was assigned to be the actor. If six-ten was picked the participant was the observer. When participants were tested separately the researcher alternated between assigning them as an actor and as an observer in order to assure equal numbers of actors and observers. For example the first participant individually tested was an observer, the second an actor, the third an observer, and so on.

Regardless of the role played, all participants were told that the purpose of the experiment was to investigate differences in recipe mixing styles between those who have cooked a great deal and those who rarely cook. They were also informed that if they were the one acting in the experiment they would simply be mixing a recipe just as they would follow a cookie recipe. If they were observing they were told to pay close attention to how the recipe was mixed and see if they would do anything differently.

The recipe was then mixed by the actor while the observer watched. The recipe consisted of six products--baking powder, baking soda, chocolate chips, flour, salt, and sugar. There were four possible brands of each product that could be used. Each of the four brands was used equally often by actors and observers. The brands were assigned pseudo-randomly prior to the participant's arrival by assigning each brand a number between one and four and then drawing a number for each product type.

Following the recipe mixing participants were asked to read a short article on culinary therapy as a brief distraction task. Upon completion of the article, or after five minutes, participants were either asked to fill out a questionnaire immediately or to return in one week at which time they were administered the questionnaire. Participants completed the questionnaire independently, whether they were tested in pairs or not. The amount of delay (about five minutes or one week) prior to filling out the questionnaire was determined such that alternating pairs of ten received alternating delays (the first ten pairs on a five minute delay, the second ten on a one week delay).

The questionnaire consisted of 6 recognition questions asking them to identify which brand they used of each product. All four possible brands were listed for each product. In addition, the most familiar brand of each product (i.e., Arm and Hammer® baking soda or Gold Medal® flour) were also given as possible choices, though none of the most recognized products (as determined by Krug & Weaver, 2005) were used in any of the mixing sessions. (Following Krug and Weaver, 2005, positive responses to these most familiar products will be referred to as familiar false alarms, FFAs.) Additionally, they were asked to rate their confidence in each identification made, choosing from a scale as follows--0%, 1-20%, 21-40%, 41-60%, 61-80%, and 81-100% confidence. Beyond these questions on product brand, the questionnaire also included questions asking them to identify themselves as the recipe mixer or observer, their level of cooking experience, and if they were the observer if there was anything they would have done different in the recipe mixing.

Results

Two-by-two ANOVAs were done to compare differences in performances as a function of the independent variables, involvement level (whether participants were actors or observers) and time delay (the amount of time between the completion of the mixing of the recipe and administration of the questionnaire, either five minutes or one week). Self-reported experience level was analyzed in a separate two-by-two ANOCOVA, with experience level added as a covariate in analyses. The participants could rate themselves as having baked cookies or a meal 0, 1, 2-5, 6-10, 11-25, or more than 25 times. These levels of baking frequency were subsequently assigned scores of 1-6, respectively, for analysis of covariance.

The dependent variables considered were accuracy (or hits), FFAs, and confidence ratings. Hits were defined as the participants selecting the product brand that they had actually used. Following each choice of product brands the participants were asked to rate their confidence in their decision: 0, 20, 40, 60, 80, or 100% confidence. In addition to an ANOVA test to analyze differences in confidence between the four groups, the confidence-accuracy relationship was also analyzed by plotting calibration curves (as a measure of absolute accuracy) and computing gamma correlations (as a measure of relative accuracy). In calibration curves, predicted performance (subjective confidence) is shown along the x-axis, whereas actual performance is shown on the y-axis. Perfect calibration is indicated by the main diagonal: items selected with 60% confidence should be identified 60% of the time, for example. Over- and under-confidence are indicated by the relative position of the points on the curve, overconfidence falling below the perfect calibration diagonal and underconfidence above.

When computing relative metacognitive accuracy using gamma correlations, each confidence/accuracy pair is compared to all the other pairs in order to assess if accuracy increases with confidence (Nelson, 1984, 1996). A score between 1 and -1 is computed for each participant such that 1.0 indicates a perfect positive correlation and -1.0 indicates a perfect negative correlation, with 0 indicating no correlation [1]. An alpha level of

$p < .05$ was adopted for all tests of statistical reliability, unless otherwise noted.

Accuracy

Mean levels of hits (correct responses) and FFAs are shown in Table 1. Those participating as observers had a slightly but not reliably higher rate of both hits and FFAs than did actors, for hits, $F(1, 596) = 1.15$; for FFAs, $F(1, 596) = .284$. The interaction between time delay and involvement was also non-significant.

The effects of the time delay on both hit rates and FFA rates were reliable. The five minute delay group had a higher hit rate (M= .36, [.03]) than the one week delay group (M= .25, [.03]), $F(1, 596) = 8.15$. The 5 minute delay group also had a lower FFA rate (M= .24, [.02]) than the one week delay group (M= .38, [.03]), $F(1, 596) = 13.92$. Again, the interaction between time delay and involvement level was not significant. Finally, after a one-week delay, participants were more likely to identify familiar products falsely (M = .37) than correctly recognize the products they actually used (M = .25).

When analyzed as a covariate, experience did not accurately predict hit rates; those participants with more experience did not perform differently from those with less experience, $F(1, 595) = .06$, nor did they differ when making FFAs, $F = (1, 595) = .89$. Curiously, participants with the greatest accuracy were those who indicated that they had baked only one time (M=.42, [.08]).

Confidence

Mean levels of confidence are also shown in Table 1. Confidence was also analyzed by a two-by-two ANOVA, with time delay and involvement as independent variables. Observers were the most confident in their choices of product brands, despite the fact that they never personally used the brands in the experiment. In fact, the difference between actors (M=44.27, [1.78]) and observers (M=53.47, [1.83]) was reliable $F(1, 596) = 12.92$. The difference in confidence between five minute delay participants (M=50.07, [1.89]) and the one week delay group

(M=47.67, [1.77]), however, was not reliable, F (1, 596) = .88, although as previously discussed, accuracy did decline over the one-week interval (overall accuracy declined from .36 to .25).

When ratings were broken down by experience level, confidence ratings did differ significantly, F (1, 595) = 6.87. In general, participants with more experience displayed greater confidence in their decisions than those with less experience: those who had baked more than 25 times had significantly greater confidence (M=52.56, [1.69]) than those participants with no baking experience (M=41.67, [7.57]).

Confidence-Accuracy Relationship

This lack of a correlation between confidence and accuracy becomes apparent when examining the calibration curves, as can be seen in Figure 1, as well as Gamma correlations, shown in Table 1. The actors in the one-week delay group were the most overconfident while actors in the five minute delay category were closer to a perfect calibration, though still overconfident. As seen by the mean Gamma scores in Table 1, the observers in both the five minute and one week delays displayed a higher correlation between confidence and accuracy than did the actors. The difference in Gamma scores between involvement levels (actors versus observers) was reliable, F (1, 76) = 6.69. In addition, the difference between the two time delays was also significant, with the five minute delay group showing a higher correlation F (1, 76) = 7.12. Once again, there was not a reliable interaction between time delay and involvement level.

DISCUSSION

Contrary to expectations, actors and observers did not differ in their accuracy; in fact, the trend was such that observers were more accurate than actors. This may be a result of increased accuracy in the observer group due to the opportunity given them to read all the product labels during the experiment while the actor mixed the recipe. In most real-world situations, non-actors are not simply observing—they are doing other activities themselves. Biederman et al. (1998) reported that electricians, carpenters, and brick masons were generally unable to recall what products they were exposed to, but that workers in asbestos factories and insulators were able to identify without help the products they used.

The finding of poorer product recognition after one week is in agreement with prior research (Krug & Weaver, 2005; Terrell & Weaver, 2005). The fact that information is retained after brief intervals indicates that poor memory displayed at one week intervals is not entirely an effect of poor encoding. Some information is encoded initially, but this information is lost during the passage of time, yet another example of what Schacter (1999) has called “transience,” or Ebbinghausian forgetting (Ebbinghaus, 1913/1885; Wixted, 2004).

Large effects of familiarity were seen in the present research, what Krug and Weaver (2005) referred to as a “familiarity bias.” After one week, participants were much more likely to identify the product with which they were familiar than the product they actually used. This further demonstrates the reconstructive nature of eyewitness memory, especially after delays; a witness will often mistake a sense of familiarity with true recollection. In fact, a single, rapid exposure to a face can increase the likelihood that the face will later be misidentified (Jacoby, Kelley, Brown, & Jasechko, 1989; Jacoby, Woloshyn, & Kelley, 1989). Furthermore, neurological studies have suggested that the neural structures mediating true recollection are different from those regions mediating familiarity-based recognition (Ranganath et al., 2004).

Although confident witnesses are often more convincing to a jury, the relationship between confidence and accuracy in eyewitness memory is modest at best at best (Kebbell & Wagstaff, 1997; Penrod & Cutler, 1995; Perfect & Hollins, 1999; Read, Lindsay, & Nicholls, 1998; Winningham & Weaver, 2000). In addition, factors influencing eyewitness confidence and eyewitness accuracy are not necessarily the same (Wells, Olson, & Charman, 2002). Consistent with previous research, the confidence-accuracy relationship was stronger after a five-minute than a one week delay. Furthermore, though accuracy declined significantly over the delay, subjective confidence did not. As is often the case in eyewitness identification, we see little evidence that confident witnesses in our research were more accurate. As a further illustration of the poor confidence-accuracy relationships, participants with greater experience had reliably higher confidence than did those with less experience. The more experienced participants were not, however, more accurate.

These results demonstrate poor accuracy in identifying specific brands of products used, even with relatively short delays. This was true even if an individual personally used a product, rather than simply observing someone else use the products. Our findings suggest that simple handling of a particular product is insufficient to induce encoding of the names of those products. In this manner, our results mirror the well-known findings of Nickerson and Adams (1982) regarding memory for a common object, in their case a penny. Although handling of pennies is common, fewer then 10% of those studied could correctly locate the features on a penny, presumably because those details are not important to how a penny was used at the time. Likewise, the importance of the brands of the products used in our study was not apparent at the time of use, reducing the likelihood of encoding.

Finally, in real-world situations, retention intervals are likely to be much longer than those used here. Therefore, while the pattern of forgetting is likely to be similar at even longer delays, the magnitude of the forgetting is likely to be increased. Just as researchers urge caution when evaluating the reliability of eyewitnesses in criminal justice settings, similar caution is warranted when evaluating the reliability of eyewitnesses offering product identification testimony following long intervals, regardless of confidence expressed by those witnesses.

REFERENCES

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25*, 475-491.

Biederman, C., Korosec, T., Lyons, J., & Williams, P. (1998, Aug 13, 1998). Toxic justice. *Dallas Observer*.

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.

Davies, G., Davies, G., & Alonso-Quecuty, M. (1997). Cultural factors in the recall of a witnessed event. *Memory, 5*, 601.

Ebbinghaus, H. (1913/1885). *Memory: A contribution to experimental psychology* (H. A. Roger & C. E. Bussenius, Trans.). New York: Teachers College.

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.

Ihlebaek, C., Love, T., Eilertsen, D. E., & Magnussen, S. (2003). Memory for a staged criminal event witnessed live and on video. *Memory, 11*, 319.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56*, 326-338.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118*, 115-125.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Kassin, S. M., & Kassin, S. M. (1984). Eyewitness identification: Victims versus bystanders. *Journal of Applied Social Psychology, 14*, 519.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research. *American Psychologist, 56*, 405-416.

Kebbell, M. R., & Wagstaff, G. F. (1997). An investigation into the influence of hypnosis on the confidence and accuracy of eyewitness recall. *Contemporary Hypnosis, 14*, 157-166.

Krug, K. S., & Weaver, C. A., III. (2005). Eyewitness memory and metamemory in product identification: Evidence for familiarity biases. *Journal of General Psychology, 132*, 429-445.

Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science, 15*, 149-154.

Loftus, E. F. (1996). *Eyewitness testimony* (2 ed.). Cambridge, MA: Harvard University Press.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109-133.

Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. *Applied Cognitive Psychology, 10*, 257-260.

Nickerson, R. S., & Adams, M. J. (1982). Long-term memory for a common object. In U. Neisser (Ed.), *Memory Observed* (pp. 163-175). San Francisco: Freeman.

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1*, 817-845.

Perfect, T. J., & Hollins, T. S. (1999). Feeling-of-knowing judgments do not predict subsequent recognition performance for eyewitness memory. *Journal of Experimental Psychology: Applied, 5*, 250-264.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia, 42*, 2-13.

Read, J. D., Lindsay, D. S., & Nicholls, T. (1998). The relation between confidence and accuracy in eyewitness identification studies: Is the conclusion changing? In C. P. Thompson (Ed.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 107-130). Mahwah, NJ, US: Lawrence Erlbaum Associates Inc Publishers.

Ross, D. F., Ceci, S. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and lineup identification: Toward a memory blending approach. In J. D. Read & D. F. Ross (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 80 100). New York, NY: Cambridge University Press.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist, 54*, 182-203.

Schmechel, R. P., O'Toole, T. P., Easterly, C., & Loftus, E. F. (2006). "Beyond the ken?" Testing jurors' understanding of eyewitness reliability evidence. *Jurimetrics, 46*, 177-214.

Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory and Cognition, 27*, 538-552.

Sherman, J. W., Groom, C. J., Ehrenberg, K., & Klauer, K. C. (2003). Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Social Cognition, 21*, 213-246.

Terrance, C. A., 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.

Terrell, T., & Weaver, C. A., III. (2005, November). *Misinformation effects in product identification testimony.* Paper presented at the 46th annual meeting of the Psychonomic Society, Toronto.

Terrell, T., & Weaver, C. A., III. (2006, November). *"Refreshing recollection" of eyewitnesses: Memory retrieval or memory creation?* Paper presented at the 47th Annual Meeting of the Psychonomic Society, Houston.

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.

Weaver, C. A. I. (2004). *A primer on eyewitness memory in product identification cases.* Paper presented at the Asbestos Allocation Conference: Apportionment of Liability in Asbestos Litigation, San Francisco.

Weaver, C. A. I., Terrell, J. T., & Holmes, A. E. (2006). Evaluating the reliability of eyewitness memory in product identification cases, *Asbestos Litigation in the 21st Century*. New Orleans: American Law Institute/American Bar Association.

Wells, G. L., Olson, E. A., & Charman, S. D. (2002). The confidence of eyewitnesses in their identification from lineups. *Current Directions in Psychological Science, 11*, 151-154.

Winningham, R. G., & Weaver, C. A., III. (2000). The effects of pressure to report more details on memories of an eyewitness event. *European Journal of Cognitive Psychology, 12*, 271-282.

Wixted, J. T. (2004). On Common Ground: Jost's (1897) Law of Forgetting and Ribot's (1881) Law of Retrograde Amnesia. *Psychological Review, 111,* 864-879.

Woolnough, P. S., & MacLeod, M. D. (2001). Watching the birdie watching you: Eyewitness memory for actions using CCTV recording of actual crimes. *Applied Cognitive Psychology, 15,* 395-411.

Yuille, J. C., Davies, G., Gibling, F., & Marxsen, D. (1994). Eyewitness memory of police trainees for realistic role plays. *Journal of Applied Psychology, 79*, 931-936.

Received: May 2006
Accepted: October 2006

Suggested Citation:

Colby, M. A. & Weaver, C. A. (2006). Comparing Eyewitness Memory and Confidence for Actors and Observers in Product Identification Situations: Extending Findings and Methodology from Criminal Justice [Electronic Version]. *Applied Psychology in Criminal Justice, 2*(2), 145-162.

Table 1.
Means of Hits, Familiar False Alarms (FFA), Confidence ratings (Conf.), and Gamma scores. (SEMs are shown in parenthesis.)

Involvement	Five-Minute Delay	One-Week Delay
Actors	Hit: 0.34 (0.04) FFA: 0.25 (0.04) Conf.: 46.00 (2.51) Gamma: 0.39 (0.15)	Hit: 0.23 (0.03) FFA: 0.35 (0.04) Conf.: 42.53 (2.53) Gamma: -0.05 (0.10)
Observers	Hit: 0.37 (0.04) FFA: 0.23 (0.03) Conf.: 54.13 (2.78) Gamma: 0.75 (0.34)	Hit: 0.27 (0.04) FFA: 0.41 (0.04) Conf.: 52.80 (2.40) Gamma: 0.41 (0.17)

ENDNOTES

[1] 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: $P = 0.5 + 0.5G$.

Figure 1

Confidence calibration curves for all 4 experimental groups. Perfect calibration is noted by the main diagonal.