

THE RELATIONSHIP BETWEEN CONFIDENCE AND ACCURACY: CURRENT THOUGHTS OF THE LIT- ERATURE AND A NEW AREA OF RE- SEARCH

Kevin Krug

Southeastern Oklahoma State University

Many believe the levels of confidence eyewitness' express when identifying criminal suspects in lineups or testifying in trials make good predictors of their memory accuracy. Traditionally known as the confidence-accuracy (CA) relationship, the assumption is that as one's confidence increases so does their level of accuracy. The research literature has addressed the CA relationship along three main lines: examining rates of confidence and accuracy in memory for general knowledge (factual information), determining if the CA relationship can be divided into subsections in which performance levels are consistent, and developing measures to raise the value of the CA relationship. The literature outlining the role of the CA relationship in criminal suspect identification is indeed extensive, but there is little mention of a new field of interest in which the CA relationship is applied to eyewitness memory for product brand names.

While there are a number of factors present during every trial which may ultimately decide the fate of a defendant, some factors are given more credence than others, particularly by the jury deciding the case. In some cases, the only available evidence to a jury may be the testimony of a witness who has claimed to have seen or been aware of the offenses committed by the defendant. Frequently, the jury uses a witness's level of subjective confi-

Correspondence concerning this article should be addressed to Dr. Kevin Krug, Department of Behavioral Sciences, Southeastern Oklahoma State University, 1405 N. Fourth Ave., PMB 4116, Durant, OK 74701-0609, E-mail: kkrug@sosu.edu

dence as evidence of his or her knowledge, a phenomenon known as meta-memory. Meta-memory judgments are made when jurors use an eyewitness's outward displays of confidence (i.e., witness uses terminology such as "absolutely sure" or "more than positive") when deciding the accuracy and relevance of that person's testimony. More specifically, metamemory is the monitoring, predicting, and controlling of one's memory. There are several ways to quantify metamemory, including judgments of learning, feelings of knowing, or confidence in the accuracies of memory (Benjamin, Bjork, & Schwartz, 1998; Nelson, Narens, & Dunlosky, 2004). The accuracy of these metamemory judgments can be detected by examining the relationship between a person's predicted and actual performance; this is known as the confidence-accuracy (CA) relationship. In some tests, these predictions are generally accurate, but in others the correlations are small or even nonexistent (Bothwell, Deffenbacher, & Brigham, 1987).

A number of court officials, including judges, believe a witness's level of subjective confidence is a good indication of his or her testimonial accuracy (Wise & Safer, 2004). A survey conducted by Brigham and Wolfskeil (1983) found that 73% of law enforcement officers, 75% of prosecutors, and 40% of defense attorneys consider witness confidence and testimonial accuracy positively correlated. Juries also consider confidence to be a good indicator of a witness' testimonial accuracy (Brigham, 1990; Brigham & Bothwell, 1983; Penrod & Cutler, 1995). In fact, one important criteria influencing a juror's opinion regarding eyewitness testimony credibility is the level of confidence shown by the witness during questioning (Bradfield & Wells, 2000; Cutler, Penrod, & Dexter, 1990; Cutler, Penrod, & Stuve, 1988; Leippe, Manion, & Romanczyk, 1992; Sporer, Penrod, Read, & Cutler, 1995). Juries are more likely to believe witnesses who appear very confident and excuse inaccuracies in their testimony compared to witnesses who appear less confident but give accurate testimony (Brewer & Burke, 2002; Lindsay, Wells, & Rumpel, 1981; Wells, Ferguson, & Lindsay, 1981). These findings are not limited solely to the United States. The citizens in countries such

as Australia, Canada, and England also believe the level of witness confidence should be weighted heavily when determining testimonial accuracy (McConkey & Roche, 1989; Noon & Hollin, 1987; Yarmey & Jones, 1983).

Even the United States Supreme Court considers an eyewitness's level of confidence a useful predictor of his or her testimonial accuracy. As a result of *Neil vs. Biggers* (1972), the Supreme Court acknowledged five criteria that should be followed when evaluating the reliability of perpetrator identification made by a witness: the amount of attention paid to the perpetrator by the witness, the accuracy of the witness's description of the perpetrator, the witness's view of the perpetrator during the crime, how certain the witness is of his or her identification, and the amount of elapsed time between the event and perpetrator identification. Wells and Murray (1983) have pointed out that with the exception of the last criterion, all other points are dependent on the witness' memory and cannot be completely verified.

Despite the belief of those in the court system and public regarding the CA relationship, the majority of the research asserts that confidence is a poor indicator of memory accuracy (Berger & Herringer, 1991; Hollin, 1981; Penrod & Cutler, 1995; Smith, Kassin, & Ellsworth, 1989; Sporer et al., 1995; Tomes & Katz, 1997). After reviewing 31 studies examining the CA relationship for eyewitness memory, Wells and Murray (1984) found the average correlation to be only $r = .07$. A more well-known study by Bothwell et al. (1987) found an average CA relationship of just $r = .25$ among a total of 35 research studies. Similar results were described by Penrod, Loftus, and Winkler (1982) who, after reviewing 16 eyewitness memory studies, found an average CA relationship of $r = .25$. Furthermore, in a recent survey of researchers who give expert testimony regarding eyewitness memory, 73% said they would testify that a witness's level of subjective confidence is not a good indicator of his or her true memory accuracy (Kassin, Tubb, Hosch, & Memon, 2001). It remains unclear, however, if this type of expert testimony decreases (Cutler, Dexter, & Penrod, 1989; Cutler, Penrod, & Dexter, 1989) or increases

(Fox & Walters, 1986) the susceptibility of the jury to rely on levels of confidence when evaluating testimony.

Not all studies examining the CA relationship for eyewitness memory have found that confidence is a poor indicator of accuracy. Several studies report high CA relationships in the neighborhood of $r = .42$ to $.63$ (Brewer, Keast, & Rishworth, 2002; Lindsay, Nilsen, & Read, 2000; Lindsay, Read, & Sharma, 1998; Stephenson, Brandstatter, & Wagner, 1983). These researchers go on to explain that confidence can be a useful predictor of accuracy once the correct statistical analyses and situational variables are taken into account.

The value of the CA relationship does appear to fluctuate depending on the mathematical methods used to assess it, domain tested (i.e., memory for general knowledge or eyewitness memory), underlying components, and manipulations engineered to improve accuracy. Once researchers are able to determine how each of these variables influence or control the CA relationship, the opportunity to build a theoretical framework capable of predicting a clear and reliable image of how confidence shadows accuracy may develop.

MATHEMATICAL ANALYSIS OF THE CA RELATIONSHIP

There are four mathematical methods used to verify the accuracy of the CA relationship that appear regularly in the research literature: the calibration curve, over/underconfidence (O/U) statistic, point biserial correlation, and the gamma statistic. These methods focus on analyzing two types of monitoring, absolute and relative. Absolute monitoring tells researchers how over or underconfident a participant is in terms of units (i.e., percentages) (Nelson, 1996). For example, a participant can be 20% overconfident or 10% underconfident. The calibration curve and over/underconfidence statistic are both measures of absolute monitoring. In contrast, relative monitoring tells us only if a participant is more or less confident of one statement compared to another

statement, and involves statements of performance probability (Nelson, 1996). The point biserial correlation and gamma statistic measure relative monitoring. Measures of absolute and relative monitoring should be included together because they both offer distinct, yet complementary, views of the CA relationship.

Calibration Curves

The calibration curve is a graph where a participant's subjective level of being correct (i.e., confidence) is plotted against his or her percentage of accurate decisions or identifications. A main diagonal line travels from the point of origin to intersect at the point where a level of 100% subjective confidence and actual performance would meet. If the participant indicated a subjective confidence level of 80% and was accurate 80% of the time, then he or she would be deemed well calibrated. If the participant gave a confidence level of 70%, but was only accurate 50% of the time, he or she would be described as overconfident. An underconfident person would demonstrate the opposite trend. Typically, absolute calibration would be indicated by data curves forming close to the main diagonal line while relative calibration is implied by an increasing line of data—as confidence levels increase so should actual performance (Lichtenstein & Fischhoff, 1977).

Several authors have used calibration curves as a data presentation method (Brewer et al., 2002; Juslin, Olsson, & Winman, 1996; Krug & Weaver, 2005; Olsson & Juslin, 1999); still, many researchers fail to collect calibration data because it is generally recommended that large participant samples, specifically a minimum of 200 data points, be used to insure stable calibration curves. In an exception, Weber and Brewer (2003) were successful in calculating calibration curves based on just 48 research subjects by requiring each to make multiple eyewitness identifications. As with most eyewitness research, the majority of calibration research concludes that participants are overconfident in judgments of their memory accuracy (Granhag, Stromwall, & Allwood, 2000; Krug & Weaver, 2005; Olsson, 2000).

Over/under-Confidence Statistic

Another type of absolute measurement is the over/under-confident statistic. This statistic has a range of -1 to 1 with perfect calibration demonstrated by a score of 0 . A participant with a score of -1 would be classified as under-confident, while a score of 1 would indicate an overconfident person. The over/under-confidence statistic is often used as a secondary analysis after computing calibration curves because it gives a more refined description of the relationship's direction (Bornstein & Zickafoose, 1999).

Point Bi-serial Correlation

The point bi-serial correlation uses a percentage category scale such as 0% , 20% . . . 100% to measure confidence, although accuracy is stated as either correct or incorrect. Once confidence levels are assessed, the point bi-serial correlates the identifications that are accurate with those that are inaccurate. The point bi-serial correlation, unlike the calibration curve and over/under-confidence statistic, does not indicate if a participant is over or under-confident. Instead, it produces both between and within-subjects correlations. The between-subjects correlation describes the CA relationship for the performance of all participants on one specific item. This is useful for situations in which two witnesses, testifying on the same subject or answering the same questions, differ in the levels of confidence they assign to their testimony or answers (Olsson & Juslin, 2002). The within-subjects correlation is appropriate for examining the CA relationship of one participant across several different items. This correlation examines a participant's accuracy regarding answers that were assigned higher levels of confidence versus answers given lower confidence levels (Olsson & Juslin, 2002).

Several researchers have proposed that the dual nature of the point biserial correlation may be the reason why the CA relationship seems to fluctuate so frequently among the reported results (Gruneberg & Sykes, 1983; Juslin et al., 1996; Perfect, Watson, & Wagstaff, 1993). In other words, researchers are not distinguishing the between-subjects from the within-subjects correla-

tions when reporting the CA relationships from their studies. Although Olsson (2000) stated one should use between-subjects correlations and that most of the reported CA relationships in the literature do appear to be between-subjects correlations, within-subjects correlations tend to be higher (Bothwell et al., 1987; Perfect & Hollins, 1997; Perfect et al., 1993; Robinson & Johnson, 1996; Smith et al., 1989). For example, while Robinson and Johnson (1996) noted within-subjects correlations between .49 and .63, none were significant. Their between-subjects correlations, on the other hand, were lower, ranging from .29 to .39, but were statistically significant (Robinson & Johnson, 1996). Other studies have also found nonsignificant within-subjects correlations but significant CA relationship differences between between- and within-subject correlations (Deffenbacher, Leu, & Brown, 1981; Gwyer & Clifford, 1997; Kebbell, Wagstaff, & Covey, 1996; Robinson & Johnson, 1996; Smith et al., 1989).

The Gamma Statistic

The gamma statistic is used to measure a type of relative monitoring known as resolution. Resolution distinguishes accurate from inaccurate identifications by having participants assign specific levels of confidence for each individual item, without regard to their overall level of confidence (Nelson, 1984). Participants can demonstrate good resolution by assigning high confidence levels to accurate identifications and low confidence levels to inaccurate choices. Unlike other correlation coefficients, the gamma correlation is used to interpret probability and not variance (Nelson, 1984, 1996). It is similar to other correlation coefficients, however, because it is interpreted on a range from +1.0 (perfect positive correlation) to -1.0 (perfect negative correlation) (Nelson, 1984).

The gamma correlation has been described as the most appropriate measure for determining resolution (Nelson, 1984). Confidence is deemed positively related to accuracy if it is greater among accurate than inaccurate identifications. Glenberg and Epstein (1987) suggest that one should exercise caution when using the gamma correlation because of the possibility of unstable

gamma scores. In the study, participants who were most confident were also most accurate, but they also displayed high levels of confidence when inaccurate.

It is recommended that both absolute and relative monitoring be used to evaluate the CA relationship. Several studies have shown that participants may demonstrate good absolute monitoring but poor relative monitoring and vice versa (Brewer et al., 2002; Brigham, 1990; Juslin et al., 1996; Olsson, 2000). Absolute and relative monitoring are essentially assessing two different aspects of a witness's confidence and accuracy, therefore a better picture of accuracy will be gained by including both evaluation methods.

THE CA RELATIONSHIP IN OTHER DOMAINS

Another domain where the magnitude of the CA relationship has been investigated is memory for general knowledge; for instance, memory for geography, text reading, or sports trivia (Keren, 1987; Koriat, Lichtenstein, & Fischhoff, 1980; Stephenson, 1984; Stephenson et al., 1983). Many researchers who have questioned individual's memory for general knowledge have found a strong CA relationship (Barclay & Wellman, 1986; Perfect, 2004; Schneider & Laurion, 1993; Tomassini et al., 1982). For example, Keren (1987) found that professional bridge players had better calibration than nonprofessionals, who reported overconfidence in their abilities. Similar calibration curves have been found in studies examining the relationship between levels of subjective confidence and the recall of details after reading text (Stephenson, 1984; Stephenson et al., 1983). Many researchers have proposed that participants tend to give better CA relationships for memory of general and text-based knowledge than for eyewitness events because they are able to use past experiences as benchmarks for assessing memory strengths and weaknesses (Hollins & Perfect, 1997; Juslin, 1994; Perfect, Hollins, & Hunt, 2000; Perfect et al., 1993). It is often the case that when people are called to testify as witnesses, they have little, if any, prior experience with eyewitness memory; therefore, it becomes difficult for them to gauge their meta-memory accuracy.

There are other studies that have failed to find the CA relationship to be stronger for one's memory of general knowledge (Griffin & Tversky, 1992; Koriat et al., 1980; Sniezek, Paese, & Switzer, 1990). Several investigations of text-based recall have found that participants indicate with high confidence ratings that they both understood and could recall text details, yet performed poorly during testing (Epstein, Glenberg, & Bradley, 1984; Glenberg, Sanocki, et al., 1987; Glenberg, Wilkinson, & Epstein, 1982).

The nature of the CA relationship in a classroom setting has also received some attention (Lundeberg, Fox, & Puncochar, 1994; Puncochar & Fox, 2004). This line of research has demonstrated that confidence levels tend to be poor indicators of accuracy, as students often assign equal levels to both their accurate and inaccurate answers. While students exhibit greater accuracy when they work in groups, they assign greater confidence levels to their inaccurate answers compared to students working separately, a phenomenon nicknamed the "two heads are worse than one" effect by Puncochar and Fox (2004). This trend continued despite attempts by the authors to reduce the high confidence ratings assigned to the inaccurate answers by providing quick feedback, additional classroom assignments, and metamemory lectures detailing the poor match between memory performance and subjective confidence. Allwood, Granhag, and Johansson (2003) applied a similar approach to an eyewitness memory experiment in which some witnesses collaborated together in pairs when assigning confidence ratings. These witnesses had better calibration and lower overconfidence when compared to witnesses who made their decisions individually (Allwood et al., 2003).

General knowledge and eyewitness memory are considered separate types of memory systems, but similar CA relationships found in both areas have led some authors to propose that they may be dependent on each other, known as cross-domain stability (Bothwell et al., 1987; Payne, Bettman, & Johnson, 1988; Smith et al., 1989; West & Stanovich, 1997). In other words, if a participant has good metamemory awareness of his or her memory

for facts, then a similar awareness of personal memory may be likely as well. Although participants tend to show overconfidence in both memory systems, it is normally stable with accurate identifications receiving higher confidence levels than those that are inaccurate (Bornstein & Zickafoose, 1999; West & Stanovich, 1997). Furthermore, within-subjects correlations are similar for both general knowledge and eyewitness memory (Hollins & Perfect, 1997; Perfect & Hollins, 1997). Bornstein and Zickafoose (1999) found similar calibration curves, over/underconfidence statistic scores, and resolution scores among general knowledge and eyewitness memory identifications. Additional support for cross-domain stability comes from findings where both easy general knowledge and eyewitness memory questions have higher CA relationships than more difficult general knowledge and eyewitness memory questions (Kebbell & Giles, 2000; Lichtenstein & Fischhoff, 1977; Wheatcroft, Wagstaff, & Kebbell, 2004). Together, these data seem to indicate the existence of a shared cognitive mechanism that is capable of controlling both general knowledge and eyewitness memory (Bornstein & Zickafoose, 1999).

It is unclear if knowledge of cross-domain stability would be of any benefit to a jury. Bornstein & Zickafoose (1999) proposed that if eyewitnesses are able to demonstrate good metamemory access to their semantic memory, then juries could also presume credible episodic memories. In theory, this would help juries determine which witnesses are likely to give credible testimony. Several issues must be addressed prior to determining if cross-domain stability will be of any value to a jury in rendering a verdict. First, semantic and episodic memory systems are considered separate systems and it is unknown if a witness who has good metamemory access to semantic memory would automatically have the same level of access to his or her episodic memory. Second, it is much easier to verify the accuracy of one's semantic than episodic memory. Third, metamemory access to semantic memory is likely to improve based on experience.

UNDERLYING COMPONENTS OF THE CA RELATIONSHIP

Another approach used in the research literature to gain a better understanding of the CA relationship involves determining if it can be divided into subsections with distinct categories. These categories used by researchers include: target present and absent lineups, choosers and non-choosers, and recognition and recall memory. If participants consistently perform well in one subsection but not another, the CA relationship may be relevant to specific circumstances.

Target Present and Absent Lineups

During a typical eyewitness experiment, participants view others acting out a crime in public or watch a film depicting criminal action. Afterwards, the witness must identify the perpetrator of the crime from a lineup or set of photographs. In target present lineups, the real perpetrator is in the lineup; however, in target absent lineups, the perpetrator is not present but foils that look very similar in appearance are used instead. Many have questioned if participants can use their subjective level of confidence to determine whether they are viewing a target present or absent lineup. Several studies seem to argue against such a proposal (Memon, Hope, & Bull, 2003; Read, 1995; Sporer, 1992; Sporer et al., 1995). In a meta-analysis by Sporer et al. (1995) consisting of 30 studies using both target present and absent lineups, the overall CA relationship was only $r = .29$. In another study, participants who made identifications from target present and absent lineups often had high levels of subjective confidence in both conditions (Read, 1995). Overall, evidence indicating a fundamental difference for the CA relationship between target present and absent lineups appears weak.

Another research procedure that is similar to the target present and absent lineups is the match and mismatch method. During the match description method, the target is present, as are several similar-looking foils. This is different from the mismatch description where the target is present but others in the lineup are

vastly different in appearance. In research by Wells and Olson (2002), both the match and mismatch conditions provided high identification rates when the target was present. When the target was absent in the match condition, participants demonstrated low rates of foil identification while those in the mismatch condition continued to pick the foils at a high rate (Wells & Olson, 2002).

In order to maximize the difference between confidence ratings for target present and absent lineups, researchers have increased the amount of time the witness is exposed to the perpetrator (Memon et al., 2003; Read, 1995). Read (1995) found that when witnesses had live interaction with the perpetrator for 4 to 15 minutes they made more correct identifications in a target present lineup, yet made more false identifications in target absent lineups compared to witnesses who saw the perpetrator for only 30 to 60 seconds. In a similar study, Memon et al. (2003) had participants watch a video of a simulated crime in which they saw the perpetrator's face for either 45 seconds (long exposure) or 12 seconds (short exposure). Participants who viewed target present and absent lineups under the long exposure time-frame had better rates of accuracy than participants in the short exposure conditions. Overall, participants with the long duration of exposure gave significantly higher confidence ratings to their identifications than those with the short duration of exposure; however, their confidence ratings did not differ between accurate and inaccurate identifications. The only group of participants in which confidence was concluded to be a modest indicator of accuracy were those with a short duration of exposure in a target present lineup. These research findings resulted in the conclusion that the amount of time exposed to the perpetrator was not a true predictor of CA relationship differences across target present and absent lineups (Memon et al., 2003; Read, 1995).

The issue of CA relationship differences between target present and absent lineups may be considered moot due to ecological validity. In reality, police officers do not know if they are using a target present or absent lineup. Also, police departments may use different criteria when deciding which suspects to include in the

lineup. Some departments may require strong evidence of a suspect's guilt before including him or her in the lineup whereas other departments may assign suspects to a lineup based solely on intuition. In the end, the compilations of lineups depends on the effectiveness of the police investigation.

Choosers Versus Non-choosers

During target present and absent research situations, a participant who identifies a person as the perpetrator (regardless of the accuracy of their decision), is known as a chooser because he or she has made a positive decision (Sporer et al., 1995). This differs from the non-chooser who makes a negative decision by refusing to make an identification or rejecting the lineup completely (Sporer et al., 1995). Often, non-choosers are seen as unreliable witnesses, although some researchers propose the difference between choosers and non-choosers may be an important distinguishing variable of the CA relationship (Brigham, 1990; Fleet, Brigham, & Bothwell, 1987; Leippe, 1980; Malpass & Devine, 1981; Wells & Lindsay, 1985). The basis of this proposal stems from the idea that choosing may represent a balanced cognitive or personality style where participants can confidently distinguish between accurate and inaccurate identifications (Cutler & Penrod, 1989a). Additionally, non-choosers, through their decision not to identify a person from the lineup, may also be conveying information regarding a cognitive decision-making process that is equally important. Wells and Olson (2002) furthered this argument by advocating that researchers must distinguish between those eyewitnesses who reject a lineup or make a negative decision because the true perpetrator is "not there" from those who make the same decision only because they "don't know."

In the past, researchers have not focused on the distinction between choosers and non-choosers, and instead often collapsed across conditions or combined errors of omissions and false identification together (Brigham, 1990; Sporer et al., 1995). Higher CA relationships have been documented for choosers than non-choosers (Brigham, 1990; Fleet et al., 1987; Pigott & Brigham, 1985; Sporer, 1992, 1993; Sporer et al., 1995). Fleet et al. (1987)

found that choosers had a CA correlation of $r = .50$ while the non-choosers were only $r = .14$. Of the few studies that have used CA calibration curves to investigate differences between choosers and nonchoosers, they too have found much better relationships for eyewitnesses who made positive decisions (Brewer et al., 2002; Weber & Brewer, 2003, 2006). The calibration curves for negative decisions, on the other hand, are normally flat and almost nonexistent. Weber and Brewer (2006) did secondary point biserial correlations and found additional differences favoring participants who were choosers. Choosers possibly have more stable cognitive or personality styles.

There are other studies that failed to find such positive results for choosers (Brewer et al., 2002; Hosch, Leippe, Marchioni, & Cooper, 1984; Murray & Wells, 1982; Sporer et al., 1995). While Murray and Wells (1982) concluded that choosers and non-choosers were equally confident, Hosch et al. (1984) found non-choosers to be more confident than choosers. Moreover, Brewer et al. (2002) compared calibration curves between choosers and non-choosers and although there were no differences in calibration curves, both groups were overconfident.

In an effort to explain the differences between choosers and non-choosers, researchers have examined the influence of various cognitive and personality styles (Fleet et al., 1987; Robinson & Johnson, 1996). The self consciousness scale, which purports to measure private self consciousness, public consciousness, and social anxiety, has been considered useful when examining the CA relationship (Fleet et al., 1987; Robinson & Johnson, 1996). It was proposed that individuals who score high on these scales, especially the private self consciousness scale, may have significant insight into their own mental processes and retrieval efforts (Robinson & Johnson, 1996). Unfortunately, research has found no link between scores on the self consciousness scale and the CA relationship (Fleet et al., 1987; Robinson & Johnson, 1996). Leippe (1980) stated that after a witness makes an identification, even if it is inaccurate, confidence increases because he or she feels bound by that decision. Self-perception theory has

also been used to explain instances where witnesses will develop an attitude that they must be correct simply because they were questioned by the police or asked to testify during a trial (Bem, 1972). The majority of researchers agree that introspective access to higher order mental processes is not the result of a stable cognitive or personality style (Brewer et al., 2002; Hosch et al., 1984; Kassin, 1985; Murray & Wells, 1982; Robinson & Johnson, 1996; Sporer, 1992, 1993).

Recognition and Recall Memory

The type of memory tested, recognition or recall, appears to have an important influence on the magnitude of the CA relationship (Lindsay et al., 1998; Perfect et al., 2000; Robinson & Johnson, 1996; Smith et al., 1989). Weak CA relationships generally below $r = .20$ are common with recognition memory tests (Robinson & Johnson, 1996; Robinson, Johnson, & Herndon, 1997; Robinson, Johnson, & Robertson, 2000; Smith et al., 1989). Placing a heavy reliance on contextual cues provided by law enforcement or attorney questioning may be a likely reason why participants seem to have such limited insight into the accuracy of their recognition memory (Robinson et al., 2000). For example, confidence levels have been known to increase for both accurate and inaccurate answers if participants view a photograph similar to the target before being asked to identify the target (Chandler, 1994). In terms of old and new face recognition tests, however, Olsson, Juslin, and Winman (1998) and Cutler and Penrod (1989b) announced good calibration and a slight CA relationship ranging from $r = .20$ to $.30$. Perhaps once the correct contextual cues are used, recognition memory formats can offer satisfactory CA relationships.

The CA relationship tends to be much stronger in studies of recall memory, with typical CA correlations ranging from $r = .53$ to $.64$ (Robinson & Johnson, 1996; Robinson et al., 1997; Robinson et al., 2000; Stephenson et al., 1983; Stephenson, Clark, & Wade, 1986). In a study conducted by Robinson et al. (2000), participants watched a video of a teacher performing her daily duties. After the movie ended, participants in the recall memory

condition were asked to remember the color of the teacher's shirt, while participants in the recognition memory condition were offered response options: purple, blue, green, or yellow. The researchers found that recall memory participants were more confident in their accurate identifications and less confident in their inaccurate identifications than were the recognition memory participants. Recall participants also had higher between- and within-subjects correlations compared to recognition participants (Robinson & Johnson, 1996). The high CA correlations observed among recall participants can be taken as evidence that they may have better meta-memory insight than recognition participants.

It is unclear why there is such a difference in the CA relationship between recognition and recall memory. Robinson et al. (2000) has shown that 40% of the variance in confidence judgments for recall participants is due to accuracy alone, while only 4-8% of the variance for recognition participants is the result of accuracy. The authors theorized that recall participants must use an additional meta-memory variable such as answer retrieval speed when making confidence judgments (Robinson et al., 2000). In other words, recall participants may use the amount of effort involved in remembering information as a type of cue, which gives them an advantage over recognition participants for whom retrieval is less of a determining factor (Kelley & Lindsay, 1993; Robinson et al., 1997). Although ease of retrieval seems like a plausible explanation for differences in the CA relationship, it has not been shown to function as a significant predictor of participant confidence or accuracy levels (Robinson et al., 2000). Other variables such as memory vividness and reaction time have also been unsuccessful in predicting CA relationship differences (Brewer, Caon, Todd, & Weber, 2006; Loftus, Donders, Hoffman, & Schooler, 1989; Lovelace, 1984; Payne, Elie, Blackwell, & Neuschatz, 1996; Robinson et al., 2000).

METHODS USED TO ENHANCE THE CA RELATIONSHIP

Instead of finding a way to subdivide the CA relationship into distinct categories, other researchers have pursued methods to at-

tain higher CA relationships (Bornstein & Zickafoose, 1999; Glenberg, et al., 1987; Kassin, 1985; Luus & Wells, 1994; Mello & Fisher, 1996; Roberts & Higham, 2002; Robinson & Johnson, 1996; Robinson et al., 2000). These methods have included feedback, biased and unbiased instructions, the cognitive interview, optimality of encoding, and hypothesis disconfirmation.

Feedback

The use of feedback as a method to enhance the CA relationship has received extensive attention in the research literature (Blagrove & Akehurst, 2000; Granhag et al., 2000; Perfect et al., 2000; Semmler, Brewer & Wells, 2004; Wells, Lindsay, & Ferguson, 1979). Rates of confidence do increase each time a witness is asked the same question several times or given confirmatory feedback, known as the standard confidence inflation effect (Allwood, Knutsson, Granhag, 2006). One issue that remains unresolved is the specific way that feedback influences CA relationship calibration and resolution judgments. Sharp, Cutler, and Penrod (1988) noted that feedback improves resolution, not confidence judgments. However, feedback also has been shown to increase calibration without effecting resolution judgments (Lichtenstein & Fischhoff, 1980). Subbotin's (1996) work supports the notion that feedback improves calibration but only for general knowledge questions that are easily answered. To complicate matters further, Bornstein and Zickafoose (1999) found that feedback improved neither resolution nor calibration. These issues will likely be resolved once the correct manner in which to deliver feedback is established.

A number of studies have given participants feedback related to internal variables, such as mental response time when answering questions, as a way to improve CA relationships (Perfect et al., 2000; Robinson et al., 2000; Wells & Bradfield, 1998). Some of this research has indicated that participants use the amount of time required to give an identification as a way to determine the accuracy of their responses (Dunning & Stern, 1994; Kassin, 1985; Kassin, Rigby, & Castillo, 1991; Smith et al., 1989); however, other research does not support such an argument (Robinson

et al., 2000;). In other words, if a participant offers a quick response to a question, he or she may subconsciously rate the certainty of this answer with higher confidence than a participant who took longer to answer the question. Both accurate and inaccurate response time feedback provided by researchers does not influence the CA relationship, even when it is given to the participants quickly (Robinson et al., 2000). Also, within-subjects manipulation of reaction time feedback was unsuccessful in increasing the CA relationship (Robinson et al., 2000). These strategies suggest that feedback given to participants regarding their internal functioning appears to have little, if any, effect on the CA relationship.

In contrast, confidence ratings can be increased by providing participants with feedback dealing with external variables, such as being told that other witnesses identified the same perpetrator (Bornstein & Zickafoose, 1999; Bradfield, Wells, & Olson, 2002; Kassin, 1985; Luus & Wells, 1994; Wells & Bradfield, 1998). In another example of external feedback, participants who watched a video of themselves making identifications before they assigned confidence levels to their choices, significantly increased CA relationship levels (Kassin, 1985). Other external feedback enhancement methods, such as having the witness write a narrative before viewing the lineup or indicating the three most likely candidates before making a final choice does not seem to increase the CA relationship (Robinson & Johnson, 1996).

It is still probable that feedback, if given in the correct manner, may serve as an extra benchmark that allows participants to make better meta-memory judgments; however, the correct type of feedback and when to deliver it still remains unclear. Bornstein and Zickafoose (1999) were able to reduce rates of overconfidence among their participants after telling them that people tend to display overconfidence when tested with general knowledge questions. Wells and Bradfield (1998) mistakenly believed that asking a witness his or her level of confidence prior to the delivery of feedback, nicknamed the “confidence-prophylactic hypothesis,” would prevent memory contamination. In reality,

this only worsened the situation when witnesses unintentionally took the information provided by the authors and made it part of their testimony. There is a risk to giving feedback to witnesses during any stage of the identification process, as it is likely to become part of the basis they will use when making confidence and accuracy judgments.

Type of instruction

Another influence on the CA relationship centers on the type of instructions given to participants (biased versus unbiased) and the types of questions (confusing versus simple) asked of them. An example of biased instructions is a participant being told that the true perpetrator is one of the choices in the lineup, while unbiased instructions state the perpetrator may or may not be in the lineup. It is also common to see the presence of positively and negatively biased questions in the literature (Fleet et al., 1987). A positively biased instruction would be “we think the person who committed the robbery is in the lineup” while an example of a negatively biased instruction would be “we don’t think the person you saw is in this lineup.” Participants given biased instructions are more likely to make incorrect identifications even in target absent lineups (Kohnken & Maass, 1988; Malpass & Devine, 1981). This was especially evident in studies where a “don’t know” option was not included and the participants knew that the crime was staged. During situations where a “don’t know” option was included and the participants were led to believe the witnessed event was indeed real, they increased their number of “don’t know” responses and seemed to take longer to verbalize their answers (Kohnken & Maass, 1988; Steblay, 1997). When participants were given unbiased instructions, they tended to have lower rates of choosing in target absent lineups (Semmler et al., 2004). Additional research, however, has found no significant differences in terms of error rates when both biased and unbiased instructions were given to participants (Fleet et al., 1987). Although biased instructions can influence a participant’s level of accuracy, other factors such as whether the participant considers the witnessed event to be real or staged, also influences his or her accuracy rates.

The effects of simplified and confusing questions have also been investigated to determine their influence on the CA relationship magnitude (Kebbell & Giles, 2000; Wheatcroft, et al., 2004). In these experiments, after viewing a video, participants were asked either confusing or simple questions, and were then asked to rate the confidence of their answers (Wheatcroft, et al., 2004). The CA relationship was meager to nonexistent if participants were asked confusing questions regarding their perpetrator identifications (Kebbell & Giles, 2000; Wheatcroft et al., 2004). Participants who received simplified questions had equal levels of confidence in both their accurate and inaccurate responses (Kebbell & Giles, 2000; Wheatcroft et al., 2004). While comparing confidence levels, researchers noted that participants in the confusing question condition actually gave higher confidence levels to their inaccurate answers than participants in the simple question condition assigned to their accurate answers (Kebbell & Giles, 2000; Wheatcroft et al., 2004). These findings suggest that law enforcement officials and attorneys should use simplified language when questioning witnesses in order to obtain the highest levels of confidence and accuracy.

The Cognitive Interview

Originally developed by Fisher and Geiselman (1992), the cognitive interview is a set of instructions and procedures intended to improve the amount of information remembered by a witness. The cognitive interview contains four memory enhancing techniques: mentally reinstating the physical environment of the witnessed event and any interactions with others that occurred; instructing the witness to recall the event in several temporal orders such as from beginning to end or end to beginning; describing the event from various visual perspectives including the perpetrator's and victim's; and having the witness report as many details of the event as possible. The witness may also be asked to include drawings of the event or other nonverbal information. Recently, a new version known as the enhanced cognitive interview has been utilized. It includes additional features such as more open response questions, new techniques for developing better rapport with the witness, and ways to minimize the

number of interview interruptions so as to increase the amount of information remembered (Fisher, Geiselman, Raymond, Jurkevich, & Warhaftig, 1987; George & Clifford, 1992).

Several authors have credited the cognitive interview with an increase in the amount of accurate information remembered without any parallel rises in inaccurate information or errors (Aschermann, Mantwill, & Kohnken, 1991; Mantwill, Kohnken, & Aschermann, 1995). Geiselman, Fisher, MacKinnon, and Holland (1986) were able to raise the amount of accurate information remembered to 35% by using the cognitive interview. Furthermore, the enhanced cognitive interview raised accuracy levels to 45% (Fisher & Geiselman, 1992). This is likely due to the specific mnemonic devices used in the cognitive interview (Gwyer & Clifford, 1997).

While the cognitive interview has been positively related to the CA relationship, it is important to note that some studies have either found a negative relationship or none at all (Deffenbacher, 1980; Wells & Murray, 1984). The cognitive interview has received criticism in the research literature by some authors, particularly for its measure of “report everything” (Mantwill et al., 1995; McCauley & Fisher, 1995; Mello & Fisher, 1996; Memon & Higham, 1999; Roberts & Higham, 2002). Not only do witnesses report more correct information when prompted to report everything, they also report a greater amount of inaccurate information (Mantwill et al., 1995; Roberts & Higham, 2002). This prompting may encourage witnesses to report all of the details they can think of regardless of the level of criterion or confidence they have assigned to these memories (Higham, 2002). Furthermore, the cognitive interview has been shown to elicit only about half correct information, most of which occurred during the first phase, mentally reinstating the physical environment and any human interactions that took place (Roberts & Higham, 2002). This raises the notion that the other three phases of the cognitive interview may be unnecessary.

Some attention in the research literature has focused on com-

paring the cognitive interview with the police standard interview in order to determine which is more effective in eliciting correct information (Gwyer & Clifford, 1997; Kebbell & Wagstaff, 1997; Mello & Fisher, 1996; Memon, Holley, Milne, Kohnken, & Bull, 1994). The standard interview differs from the cognitive interview in that it contains only half the questions and limits the number of attempts to retrieve information from memory. Also, the questions from the standard interview tend to be more leading and direct in nature when compared to the cognitive interview. During a recent meta-analysis, Kohnken, Milne, Memon, and Bull (1999), found the cognitive interview consistently out-performed the standard interview in generating correct information remembered by witnesses. Mello and Fisher (1996) found that both young and old adults remembered more correct information with the cognitive interview compared to those who received the standard interview, although the cognitive interview also elicited more inaccurate statements than the standard interview. The cognitive interview does increase the CA relationship regarding person identification, while the standard interview produces a higher CA relationship for object identification (Gwyer & Clifford, 1997).

Studies that have examined the influence of the cognitive interview on recognition and recall memory have produced mixed results (Cutler, Penrod, & Martens, 1987; Sanders, 1984; Schooler & Engstler-Schooler, 1990). The contextual reinstatement is capable of improving recognition accuracy, but this is not a consistent finding (Gwyer & Clifford, 1997; Krafka & Penrod, 1985; Sporer et al., 1995). One explanation, dubbed the outshining hypothesis, claims that cues provided by the stimulus (i.e., seeing the perpetrator in the lineup) should be sufficient and the cognitive interview's contextual reinstatement mnemonic devices are either redundant or introduce interference into a witness' memory (Smith, 1988). This explanation has merit, as the cognitive interview does a better job of increasing recall, as opposed to recognition, of accurate information (Gwyer & Clifford, 1997).

The Optimality Encoding Hypothesis

Originally coined by Deffenbacher (1980), the optimality encoding hypothesis proposes that the CA relationship will be at its maximum level when conditions at encoding, storage, and retrieval are optimal. Situations that would qualify as optimal include the perpetrator not wearing a disguise, good lighting and environmental conditions, or watching the crime unfold over a lengthy period of time. Other authors have also supported the optimality hypothesis (Bothwell et al., 1987; Brigham, 1990; Clifford & Hollin, 1981; Krafka & Penrod, 1985). For example, Brigham (1990) found CA relationships ranging from $r = .38$ to $.45$ for targets that were distinctive versus $r = .25$ for non-distinctive targets. Furthermore, conditions that have low optimality tend to produce overconfidence and low calibration among participants (Olsson, 2000). It appears the conditions under which a participant witnesses a crime are critical in predicting the strength of the CA relationship.

Other researchers have suggested that the strong CA relationships initially predicted by the optimality hypothesis are instead the result of variability present during the encoding, storage, and retrieval conditions (Brewer et al., 2006; Lindsay et al., 1998; Lindsay et al., 2000). The reliance of researchers on homogenous participant groups and very controlled environmental settings actually minimizes variability, thereby restricting the size of the CA relationship. In a real world setting, the conditions at encoding, storage, and retrieval vary, so researchers should allow for the same variability in their laboratory conditions to truly gauge the size of the CA relationship.

Hypothesis disconfirmation

Another procedure used to get sounder CA relationships from witnesses is to have them justify why their identifications may be accurate or inaccurate before they render a final confidence decision (Brewer et al., 2002; Koriat et al., 1980; Sniezek et al., 1990). The methodology of these experiments is diverse. Brewer et al., (2002) had participants make their perpetrator identifications before they were told to justify their choices, whereas

Tetlock and Kim (1987) advised participants that they would have to justify their decision before asking them to make an identification. It has been proposed that requiring participants to explain their answers will result in further evaluation and scrutiny of perpetrators. The results of research examining hypothesis disconfirmation have gained some support (Cutler et al., 1987; Hoch, 1985; Robinson & Johnson, 1996), while other studies have concluded that it has no real influence (Griffin, Dunning, & Ross, 1990).

Some studies have found that hypothesis disconfirmation reduces participant levels of overconfidence (Griffin et al., 1990; Hoch, 1985; Sniezek et al., 1990). For instance, in a study by Brewer et al. (2002) the participants in the hypothesis disconfirmation condition made fewer positive identifications that were rated as highly confident compared to those participants who were not told to justify their identifications. Also, the CA relationship in the hypothesis disconfirmation conditions were found to be higher and better calibrated compared to those participants who were told to simply make an identification. The authors concluded that requiring participants to justify the accuracy of their identifications decreases rates of overconfidence and lowers the possibility of giving insufficient attention to the decision-making process, which encourages better meta-memory judgments of confidence and accuracy.

Other research has concluded that requiring participants to further justify why their identifications may be accurate or inaccurate tends to have no real bearing on the magnitude of the CA relationship (Cutler et al., 1987; Robinson & Johnson, 1996). Tetlock and Kim (1987) found that requiring participants to justify their identifications seemed to only benefit the magnitude of the CA relationship during cases in which they were warned prior to viewing the lineup that they would be held accountable for their decisions. Furthermore, hypothesis disconfirmation helps to increase the CA relationship for recognition but not for recall memory (Robinson & Johnson, 1996). These studies, however, have been criticized for having small participant groups and not

using calibration as a measurement tool (Brewer et al., 2002).

The research literature investigating hypothesis disconfirmation as a means to improve confidence and accuracy judgments is meager. Although the hypothesis disconfirmation approach does seem capable of improving the CA relationship, it may only be situation specific for those favoring recognition memory. Further research is necessary to determine the specific situations where hypothesis disconfirmation would benefit the CA relationship.

SUMMARY AND CONCLUSIONS

Many outside of the research community consider an eyewitness' level of subjective confidence to be a valid indicator of his or her accuracy. This is typically evident in a courtroom setting where officials and jurors tend to give the most credence to witnesses who appear very confident. Contrary to this popular belief, a person's level of subjective confidence is not a valid indicator of his or her accuracy. Most scientific studies have found the CA relationship to be relatively weak or nonexistent; in fact, this is one of the most consistent findings in the memory research literature (Tomes & Katz, 1997).

Recently, there has been an effort to increase the ecological validity of the methodology used to measure the CA relationship in order to make it more marketable in the courtroom. An example of this trend includes Brewer and Wells' (2006) suggestions to not only use fewer anchor points on confidence rating scales but to also include verbal statements (i.e., very unsure, unsure, sure, and, very sure) instead of percentages (0%, 20% . . . 100%). Another attempt made by Weber and Brewer (2003) involved having research participants choose from 2 photographs presented at the same time (pair face comparisons) instead of using the standard "six pack" simultaneous photo array favored by law enforcement personnel. Future research along similar lines will make it easier to form general guidelines that the average juror can both understand and apply during deliberation, which is one of the main goals of CA relationship research.

One area of research that has received little mention is eyewitness memory and the CA relationship in civil cases. Eyewitness memory research in civil cases brings up several issues that may not necessarily be answered by memory research regarding criminal activity or cases. First, a research methodology must be designed to determine the reliability of an eyewitness's memory for the names of products. Second, many civil plaintiffs have no real reason to remember the names of products, locations, or dates. They often have no reason to deem it necessary to remember such information as an eyewitness. This is markedly different from situations in which people witness a crime, whereby they likely realize the importance of remembering details in order to report them to law enforcement.

In a study examining eyewitness memory for product recall, Krug and Weaver (2005) instructed participants to mix together 8 common household cooking products according to a recipe. The participants returned after a time delay (5 minutes, 1 week, or 2 weeks) unaware they would be asked to identify the names of the products they used. They were either asked to identify products through a recognition (i.e., multiple choice) or recall (i.e., fill in the blank) test format. Participants rated on a scale (0, 20%, 40%, 60%, 80%, or 100%) how confident they were in their product identifications (Krug & Weaver, 2005). As a way to examine the susceptibility of eyewitnesses to familiarity, Krug and Weaver (2005) listed well known household cooking items (i.e., Gold Medal Flour and Morton Salt), defined as false alarms, as answer options on the recognition tests even though these particular items were never used in the recipe mixing. Accuracy rates were higher on the recognition than recall tests, but only at chance levels (20% versus 5%). Participants had better recognition accuracy after the 5 minute delay, but this reached only 26%. For some of the product categories (flour, pepper, salt, and sugar), the recognition participants identified the false alarms as being present at a significantly higher rate than the actual product brands. Participants showed poor calibration and no significant differences were found in the confidence levels they assigned to their accurate and inaccurate product identifications. The research by Krug and

Weaver (2005) was successful in creating a useful methodology for testing eyewitness memory in product liability cases.

This author believes research into eyewitness memory for product brand names that follows the evidence rules of a civil court format is a promising new area for the CA relationship. First, juries in civil cases must follow different rules than juries in criminal cases, in particular the preponderance of evidence rule. A civil court jury only has to be mostly certain when assigning guilt while a criminal jury must believe beyond a reasonable doubt in order to convict. Second, Krug and Weaver's (2005) participants had a difficult time remembering the names of the products they used, but there may be other product features that are easier to remember, such as texture or smell. It is important to identify such features and determine if there are any corresponding changes in the CA relationship. Third, although it would be easy to incorporate the methodology used to study memory and the CA relationship in criminal cases and simply apply it to memory for product brand names, there are other research ideas that may more aptly apply to eyewitness memory in civil cases. For instance, using the model set forth by Krug and Weaver (2005), participants should be allowed to choose the products they wanted to use rather than those assigned to them. In this case, participants may have better memory of the products they used. It would be interesting to determine what influence this would have on the CA relationship. Another research suggestion involves having other products present, but not used, during the recipe mixing to determine if these other products would contaminate the memories of the participants.

In conclusion, the research literature has devoted considerable attention to gaining a better understanding of the CA relationship by examining participant performance levels in multiple domains, deciphering whether it can be divided into distinct categories, or devising ways to enhance it. Although some success has been demonstrated by giving participants specific types of feedback or requiring them to consider why their identifications may be incorrect, little has been found to suggest that a partici-

part's level of subjective confidence is a useful predictor of his or her true accuracy. The general consensus of the research literature states that subjective confidence is not only an unstable indicator of one's meta-memory accuracy, but is also easily changed.

REFERENCES

- Allwood, C. M., Granhag, P. A., & Johansson, M. (2003). Increased realism in eyewitness judgments: The effect of dyadic collaboration. *Applied Cognitive Psychology, 17*, 545-561.
- Allwood, C. M., Knutsson, J., Granhag, P. A. (2006). Eyewitnesses under influence: How feedback affects the realism in confidence judgments. *Psychology, Crime, & Law, 12*, 25-38.
- Aschermann, E., Mantwill, M., & Kohnken, G. (1991). An independent replication of the effectiveness of the cognitive interview. *Applied Cognitive Psychology, 5*, 489-495.
- Barclay, C., R., & Wellman, H. M. (1986). Accuracies and inaccuracies in autobiographical memories. *Journal of Memory & Language, 25*, 93-103.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol 6). New York: Academic Press.
- 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.
- Berger, J. D., & Herringer, L. G. (1991). Individual differences in eyewitness recall accuracy. *Journal of Social Psychology, 131*, 807-813.
- Blagrove, M., & Akehurst, L. (2000). Effects of sleep loss on confidence-accuracy relationships for reasoning and eyewitness memory. *Journal of Experimental Psychology: Applied, 6*, 59-73.
- Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied, 5*, 76-88.
- Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*, 691-695.
- Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior, 24*, 581-594.
- Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and accuracy. *Journal of Applied Psychology, 87*, 112-120.
- Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26*, 353-364.
- Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law & Human Behavior, 30*, 31-50.
- Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8*, 44-56.

- Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12, 11-30.
- Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. *Basic and Applied Social Psychology*, 11, 101-115.
- Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law and Human Behavior*, 7, 19-30.
- Brigham, J. C., & Wolfskeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identification. *Law and Human Behavior*, 7, 337-349.
- Chandler, C. C., (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. *Memory and Cognition*, 22, 273-280.
- Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology*, 66, 364-370.
- Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1989). Expert testimony and jury decision making: An empirical analysis. *Behavioral Sciences & the Law*, 7, 215-225.
- Cutler, B. L., & Penrod, S. D. (1989a). Forensically relevant moderators of the relationship between eyewitness identification accuracy and confidence. *Journal of Applied Psychology*, 74, 650-652.
- Cutler, B. L., & Penrod, S. D., (1989b). Moderators of the confidence^accuracy correlation in face recognition: The role of information processing and base-rates. *Applied Cognitive Psychology*, 3, 95-107.
- Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1989). The eyewitness, the expert psychologist, and the jury. *Law and Human Behavior*, 13, 311-332.
- Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior*, 14, 185-191.
- Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, 72, 629-637.
- Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior*, 12, 41-55.
- Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about the relationship? *Law and Human Behavior*, 4, 243-260.
- Deffenbacher, K. A., Leu, J. R., & Brown, E. L. (1981). Memory for faces: Testing method, encoding strategy and confidence. *American Journal of Psychology*, 94, 13-26.
- Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate identifications via inquires about decision processes. *Journal of Personality and Social Psychology*, 67, 818-835.
- Epstein, W., Glenberg, A. M., & Bradley, M. M. (1984). Coactivation and comprehension: Contribution of text variables to the illusion of knowing. *Memory & Cognition*, 12, 355-360.
- Fisher, R. P. & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C Thomas.
- Fisher, R. P., Geiselman, R. E., Raymond, D. S., Jurkevich, L. M., & Warhaftig, M. L. (1987). Enhancing enhanced eyewitness memory: Refining the cognitive interview. *Journal of Police Science & Administration*, 15, 291-297.
- Fleet, M. L., Brigham, J. C., & Bothwell, R. K. (1987). The confidence-accuracy relationship: The effects of confidence assessment and choosing. *Journal of Applied Social Psychology*, 17, 171-187.

- Fox, S. G., & Walters, H. A. (1986). The impact of general versus specific expert testimony and eyewitness confidence upon mock juror judgment. *Law and Human Behavior, 10*, 215-228.
- Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *American Journal of Psychology, 99*, 385-401.
- George, R., & Clifford, B. R. (1992). Making the most of witnesses. *Policing, 8*, 185-198.
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*, 84-93.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116*, 119-136.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition, 10*, 597-602.
- Goldsmith, R. E. (1989). Reducing spurious response in a field study. *Journal of Social Psychology, 129*, 201-212.
- Granhag, P. A., Stromwall, L. A., Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology, 14*, 397-420.
- Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality and Social Psychology, 59*, 1128-1139.
- Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411-435.
- Gruneberg, M. M., & Sykes, R. N. (1993). The generalisability of confidence-accuracy studies in eyewitnessing. *Memory, 1*, 185-189.
- Gwyer, P., & Clifford, B. R. (1997). The effects of the cognitive interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology, 11*, 121-145.
- Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory and Cognition, 30*, 67-80.
- Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 719-731.
- Hollin, C. R. (1981). Nature of the witnessed incident and status of interviewer as variables influencing eyewitness recall. *British Journal of Social Psychology, 20*, 295-296.
- Hollins, T. S., & Perfect, T. J. (1997). The confidence-accuracy relation in eyewitness memory: The mixed question type effect. *Legal and Criminological Psychology, 2*, 205-218.
- Hosch, H. M., Leippe, M. R., Marchioni, P. M., & Cooper, D. S. (1984). Victimization, self-monitoring, and eyewitness identification. *Journal of Applied Psychology, 69*, 280-288.
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226-246.
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1304-1316.

- Kassin, S. M. (1985). Eyewitness identification: Retrospective self-awareness and the accuracy-confidence relationship. *Journal of Personality and Social Psychology*, *49*, 878-893.
- Kassin, S. M., Rigby, S., & Castillo, S. R. (1991). The accuracy[^]confidence correlation in eyewitness testimony: Limits and extensions of the retrospective self-awareness effect. *Journal of Personality and Social Psychology*, *61*, 698-707.
- Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the 'general acceptance' of eyewitness testimony research: A new survey of the experts. *American Psychologist*, *56*, 405-416.
- Keibell, M. R., & Giles, D. C. (2000). Some experimental influences of lawyers' complicated questions on eyewitness confidence and accuracy. *Journal of Psychology: Interdisciplinary and Applied*, *134*, 129-139.
- Keibell, 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.
- Keibell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology*, *87*, 653-662.
- 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.
- Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior & Human Decision Processes*, *39*, 98-114.
- Kohnken, G., & Maass, A. (1988). Eyewitness testimony: False alarms on biased instructions? *Journal of Applied Psychology*, *73*, 363-370.
- Kohnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime & Law*, *5*, 3-27.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 107-118.
- Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experience on eyewitness identification. *Journal of Personality and Social Psychology*, *49*, 58-69.
- Krug, K. S., & Weaver, C. A. III. (2005). Eyewitness memory and product identification: Evidence for familiarity biases. *Journal of General Psychology*, *132*, 429-445.
- Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior*, *4*, 261-274.
- Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? *Journal of Personality and Social Psychology*, *63*, 181-197.
- Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior & Human Performance*, *20*, 159-183.
- Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior & Human Performance*, *26*, 149-171.
- Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witnesses' versus investigators' confidence in the accuracy of witnesses' identification decisions. *Law and Human Behavior*, *24*, 685-697.
- Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science*, *9*, 215-218.

- 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., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition*, *17*, 607-616.
- Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 756-766.
- Lundeberg, M., Fox, P. W. & Puncoschar, J. (1994). Highly confident, but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, *86*, 114-121.
- Luus, C. A., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology*, *79*, 714-723.
- Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and absence of the offender. *Journal of Applied Psychology*, *66*, 482-489.
- Mantwill, M., Kohnken, G., & Aschermann, E. (1995). Effects of the cognitive interview on the recall of familiar and unfamiliar events. *Journal of Applied Psychology*, *80*, 68-78.
- McCauley, M., & Fisher, R. P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology*, *80*, 510-516.
- McConkey, K. M., & Roche, S. M. (1989). Knowledge of eyewitness memory. *Australian Psychologist*, *24*, 377-384.
- Mello, E. W., & Fisher, R. P. (1996). Enhancing older adult eyewitness memory with the cognitive interview. *Applied Cognitive Psychology*, *10*, 403-417.
- Memon, A., & Higham, P. A. (1999). A review of the cognitive interview. *Psychology, Crime and Law*, *5*, 177-196.
- Memon, A., Holley, A., Milne, R., Kohnken, G., & Bull, R. (1994). Towards understanding the effects on interview training in evaluating the cognitive interview. *Applied Cognitive Psychology*, *8*, 641-659.
- Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, *94*, 339-354.
- Murray, D. M., & Wells, G. L. (1982). Does knowledge that a crime was staged affect eyewitness accuracy? *Journal of Applied Social Psychology*, *12*, 42-53.
- Neil v. Biggers, 409 U.S. 188 (1972).
- 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, not of the absolute performance on an individual item. *Applied Cognitive Psychology*, *10*, 257-260.
- Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on meta-memory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods*, *9*, 53-69.
- Noon, E., & Hollin, C. R. (1987). Lay knowledge of eyewitness behaviour: A British survey *Applied Cognitive Psychology*, *1*, 143-153.
- Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *Journal of Applied Psychology*, *85*, 504-511.
- Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identifications? *Journal of Applied Psychology*, *84*, 42-49.

- Olsson, N., & Juslin, P. (2002). Calibration of confidence among eyewitnesses and earwitnesses. In P. Chambers, & M. Izaute. (Eds.), *Metacognition: Process, function and use* (pp. 203-218). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied*, 4, 101-118.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534-552.
- Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, 35, 261-285.
- Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, 1, 817-845.
- Penrod, S., Loftus, E. F., & Winkler, J. D. (1982). The reliability of eyewitness testimony: A psychological perspective. In R. Bray & N. Kerr (Eds.), *The psychology of the courtroom* (pp. 119-168). New York: Academic Press.
- Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgments in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, 18, 157-168.
- Perfect, T. J., & Hollins, T. S. (1997). The confidence-accuracy relation in eyewitness event memory: The mixed question type effect. *Legal and Criminological Psychology*, 2, 205-218.
- Perfect, T. J., Hollins, T. S., & Hunt, A. L. R. (2000). Practice and feedback effects on the confidence-accuracy relation in eyewitness memory. *Memory*, 8, 235-244.
- Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology*, 78, 144-147.
- Pigott, M. A., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology*, 70, 547-555.
- Puncochar, J. M., & Fox, P. W. (2004). Confidence in individual and group decision making: When "two heads" are worse than one. *Journal of Educational Psychology*, 96, 582-591.
- Read, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology*, 9, 91-121.
- Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied*, 8, 33-43.
- Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence-accuracy correlation. *Journal of Applied Psychology*, 81, 587-594.
- Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*, 82, 416-425.
- Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied*, 6, 207-221.

- Sanders, G. S. (1984). Effects of context cues on eyewitness identification responses. *Journal of Applied Social Psychology, 14*, 386-397.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71.
- Schneider, S. L., & Laurion, S. K. (1993). Do we know what we've learned from the news? *Memory & Cognition, 21*, 198-209.
- Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology, 89*, 334-346.
- Sharp, G. L., Cutler, B. L., & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgments. *Organizational Behavior and Human Decision Processes, 42*, 271-283.
- Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies and D. M. Thompson (Eds.), *Memory in context: Context in Memory* (pp 13-34). Chichester, England: Wiley.
- Smith, V. L., Kassin, S. M., & Ellsworth, P. C. (1989). Eyewitness accuracy and confidence: Within-versus between-subjects correlations. *Journal of Applied Psychology, 74*, 356-359.
- Sniezek, J. A., Paese, P. W., & Switzer, F. S. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46*, 264-282.
- Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times, and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22*, 157-180.
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*, 22-33.
- 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.
- Stebly, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*, 283-297.
- Stephenson, G. M. (1984). Accuracy and confidence in testimony: A critical review and some fresh evidence. In D. J. Muller, D. E. Blackman, & A. J. Chapman (Eds.), *Psychology and law: Topics from an international conference* (pp. 229-248). Chichester, England: Wiley.
- Stephenson, G. M., Brandstatter, H., & Wagner, W. (1983). An experimental study of social performance and delay on the testimonial validity of story recall. *European Journal of Social Psychology, 13*, 175-191.
- Stephenson, G. M., Clark, N. K., & Wade, G. S. (1986). Meetings make evidence? An experimental study of collaborative and individual recall of a simulated police interrogation. *Journal of Personality and Social Psychology, 50*, 1113-1122.
- Subbotin, V. (1996). Outcome feedback effects on under- and overconfident judgments (general knowledge tasks). *Organizational Behavior and Human Decision Processes, 66*, 268-276.
- Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52*, 700-709.
- Thompson, W. B., & Mason, S. E. (1996). Instability of individual differences in the association between confidence judgments and memory performance. *Memory & Cognition, 24*, 226-234.
- Tomassini, L. A., Solomon, I., Romney, M. B., & Krogstad, J. L. (1982). Calibration of auditor's probabilistic judgments: Some empirical evidence. *Organizational Behavior and Human Performance, 30*, 391-406.

- Tomes, J. L., & Katz, A. N. (1997). Habitual susceptibility to misinformation and individual differences in eyewitness memory. *Applied Cognitive Psychology, 11*, 233-251.
- Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology, 88*, 490-499.
- Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy, and response latency. *Applied Cognitive Psychology, 20*, 17-31.
- Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*, 360-376.
- Wells, G. L., Ferguson, T. J., & Lindsay, R. C. L. (1981). The tractability of eyewitness confidence and its implication for triers of fact. *Journal of Applied Psychology, 66*, 688-696.
- Wells, G. L., & Lindsay, R. C. L. (1985). Methodological notes on the accuracy-confidence relation in eyewitness identifications. *Journal of Applied Psychology, 70*, 413-419.
- Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64*, 440-448.
- Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil v. Biggers criteria for judging eyewitness accuracy? *Journal of Applied Psychology, 68*, 347-362.
- Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155-170). New York: Cambridge University Press.
- Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8*, 155-167.
- West, R. F. & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review, 4*, 387-392.
- Wheatcroft, J. M., Wagstaff, G. F., & Kebbell, M. R. (2004). The influence of courtroom questioning style on actual and perceived eyewitness confidence and accuracy. *Legal and Criminological Psychology, 9*, 83-101.
- Wise, R. A., & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology, 18*, 427-443.
- Yarmey, A. D., & Jones, H. P. T. (1983). Is the psychology of eyewitness identification a matter of common sense? In S. Lloyd-Bodstock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 13-40). Chichester, England: Wiley.

Received: June 2006

Accepted: May 2007

Suggested Citation:

Krug, K. (2007). The relationship between confidence and accuracy: Current thoughts of the literature and a new area of research [Electronic Version]. *Applied Psychology in Criminal Justice, 3*(1), 7-41.