Confidence and accuracy of eyewitness identification

James D. Sauer          Neil Brewer
University of Portsmouth Flinders University

Corresponding author: Neil Brewer
School of Psychology
Flinders University
GPO Box 2100, Adelaide
S. Aust. 5001
Australia

Email: neil.brewer@flinders.edu.au

Supported by Australian Research Council Grant DP1093210 and Flinders FRG Grant.
Case Study: In-court expressions of confidence.

In 1982, Robert Clark was convicted of rape, kidnapping and armed robbery. Twenty-three years later he was exonerated. The victim’s mistaken identification of Clark as the perpetrator was one of the primary contributing factors to Clark’s wrongful conviction (Innocence Project, 2013). In court, the victim testified “I will never forget the face, the skin color and his voice” (Benjamin N. Cardozo School of Law Yeshiva University, 2013). This testimony implies a high degree of certainty. However, a more thorough consideration of the identification evidence suggests that this level of certainty was misleading. Several days after the attack, the victim was shown a mug book of potential suspects and identified someone other than Clark. Almost a month after the attack, the victim viewed a photo lineup containing Clark’s picture, and identified Clark as someone “... looking very much like the person that committed the crimes” (Benjamin N. Cardozo School of Law Yeshiva Univeristy, 2013). Two days later the victim identified Clark from a live lineup in which Clark was the only lineup member whose photograph had also appeared in the previous photo lineup. As indicated by her testimony, by the time the witness reached court there was little doubt in her mind that Clark was her attacker.

A number of important points can be drawn from these case facts to illustrate that in-court expressions of confidence are not reliable predictors of identification accuracy. First, the level of confidence expressed in court appears inflated compared to the level of confidence that may be inferred from the statement that accompanied the victim’s initial identification of Clark. Second, the level of confidence expressed in court is inconsistent with the victim’s initial identification of someone other than Clark. The exact mechanisms for the apparent confidence inflation in this case are unclear, though we can speculate about the effects of pre-court preparation, and the implicit feedback associated with (a) seeing Clark’s face (and only Clark’s face) in two separate lineups and (b) knowing that the case against
Clark was proceeding to court. What is clear, however, is that the level of confidence expressed in court was not consistent with the victim’s prior identification behaviour, the victim’s prior expressions of confidence, or the accuracy of the victim’s identification decision. The level of certainty expressed in court was not informative about the quality of the witness’ memory at the time of the identification, or the extent to which the identified lineup member matched the witness’ memory of the culprit. Thus, it can only have undermined attempts to evaluate the reliability of the identification evidence.
Confidence and accuracy of eyewitness identification

The demonstrated fallibility of eyewitness identification evidence has motivated researchers to search for indices of identification decision accuracy. These are generally characteristics of the identification decision that are likely to be informative about the nature of the decision process and, importantly, considered to speak to the quality of the witness’ memory for the event and the culprit. For example, potential indices include the witness’ confidence in the accuracy of their identification decision and the speed with which the identification response is made. Confidence is the most researched index of identification accuracy and the focus of this chapter. Although research demonstrates that (a) confidence can provide useful diagnostic information about identification accuracy, (b) confidence influences perceptions of both witness credibility and the likely reliability of the identification evidence, and (c) consideration of witness confidence should accompany any evaluation of likely identification accuracy, formal mechanisms for measuring and interpreting identification confidence are non-existent in identification test protocols across the world.

The initial sections of this chapter discuss perceptions of the confidence-accuracy (CA) relationship held by lay-persons and decision-makers in the criminal justice system, and then review theoretical and empirical investigations of the CA relationship. Here we also highlight areas of convergence and divergence between these legal, theoretical, and empirical perspectives. Given that much of this has been covered in depth elsewhere (e.g., Brewer, 2006; Brewer & Weber, 2008; Brewer, Weber, & Semmler, 2005), a significant part of this chapter is then spent addressing a largely ignored but extremely important issue relating to eyewitness identification confidence: how current identification test practices fail to exploit the applied value of confidence as an index of identification accuracy. Finally, we discuss
how theories of recognition memory and confidence processing highlight new – and important – opportunities for using confidence in the criminal justice system.

**Perceptions of eyewitness confidence in legal settings**

Three lines of evidence converge to demonstrate the influence eyewitness confidence can exert on decision making in legal settings. First, survey research has illustrated that police, lawyers and jurors clearly believe confidence is reliably linked to accuracy (Deffenbacher & Loftus, 1982; Potter & Brewer, 1999). Second, mock-juror research – where jury-eligible samples make decisions in some form of trial simulation – has consistently shown that experimental manipulations of witness confidence affect mock-jurors’ perceptions of witness credibility and defendant guilt (Bradfield & Wells, 2000; Brewer & Burke, 2002; Cutler, Penrod & Dexter, 1990; Cutler, Penrod & Stuve, 1988; Lindsay, Wells & Rumpel, 1981). Finally, the United States Supreme Court has ruled that courts must consider confidence when evaluating the reliability of identification evidence (*Neil v. Biggers*, 1972). Importantly, although these lines of evidence demonstrate that legal decision-makers believe that witness confidence can predict identification accuracy, they say nothing about whether witness confidence actually predicts identification accuracy. Nor do they speak to decision-makers’ awareness of issues that may undermine the CA relationship. Finally, as we argue later in this chapter, beliefs about the influence confidence can exert in legal settings together with failures to appreciate the limits of the CA relationship may have (a) inadvertently obstructed the development of protocols necessary for ensuring a meaningful CA relationship in forensic settings, and (b) led to practices that (intentionally or not) obscure a vital function of eyewitness confidence.

**Theoretical perspectives on the CA relationship**

There are solid theoretical grounds for expecting a positive relationship between confidence and accuracy for recognition decisions. Current theories of confidence processing
for recognition decisions can be broadly classified into two categories: decisional and post-decisional locus models. These theories were intended to describe the mechanisms underlying decision-making and confidence processing for simple decision-making tasks. However, the mechanisms described can be extended to provide a compelling rationale for predicting a positive CA relationship for more complex decision-making tasks, such as when a witness views a lineup and must decide which, if any, of the lineup members is the culprit (at least in cases where the witness identifies a lineup member as the culprit). We first present these accounts as they relate to basic memory tasks and then extend them to account for the CA relationship for eyewitness identification decisions. An understanding of the basic processes linking response and confidence for recognition decisions is particularly important as it motivates our later discussions of (a) the importance of considering confidence when evaluating identification evidence (and, consequently, the importance of appropriately assessing witness confidence), and (b) novel uses of confidence in legal settings.

Decisional locus (or single-process) models, such as those developed within the signal detection theory (SDT) framework (e.g., Green & Swets, 1966; Macmillan & Creelman, 1991), argue that recognition decisions and confidence emerge simultaneously, and are based on the same information. According to SDT-based models, when attempting to judge whether or not a presented stimulus has been previously seen, that stimulus is compared to a representation of a previously viewed item stored in memory. This comparison generates a value which lies somewhere along a familiarity continuum, with previously viewed stimuli generating values (generally) towards the higher end of the continuum and previously unseen stimuli producing values (generally) towards the lower end of the continuum. Stimuli producing values exceeding a preset criterion are judged to be previously seen, while those resulting in values below the criterion are deemed to be previously unseen. Confidence is thought to index the strength of this familiarity value, or the distance by which the obtained
familiarity value exceeds the preset criterion. Thus, a comparison of a presented stimulus and the memorial representation of a previously viewed stimulus generates a value which, when compared to a preset criterion, simultaneously establishes the decision and decision confidence. This shared evidential basis for the decision and confidence supports a positive CA relationship.

The extension of this SDT theoretical framework to a lineup context (see also Horry, Palmer & Brewer, 2012; Palmer & Brewer, 2012; Palmer, Brewer & Weber, 2010) provides a clear basis for expecting a positive relationship between witness confidence and identification accuracy when the witness identifies a lineup member as the culprit. When viewing a lineup, let us assume that an initial inspection of the lineup members reveals one lineup member that the witness believes might be the perpetrator. The witness can then compare this individual to their memory of the offender. The extent to which this lineup member matches the witness’ memory of the offender will form the primary evidential basis for both the decision and confidence. When the degree of match between the lineup member and witness’ memory is high, confidence and accuracy are also likely to be high. As this degree of match decreases so, too, do confidence and likely accuracy.

Post-decisional locus models assume that confidence is determined after a decision is reached. However, these models also propose a direct relationship between the evidence upon which the decision and confidence are based and, therefore, predict a positive relationship between confidence and accuracy. In the context of a recognition decision, post-decisional models generally assume that a comparison of a presented stimulus with a memorial representation of a previously seen stimulus will provide evidence favouring either of two response alternatives: this stimulus is the same as one seen before or this stimulus is not the same as one seen before. Evidence favouring each of these two alternatives is stored in independent accumulators. Each accumulator has a preset criterion and a response is made
when the evidence in one of the accumulators exceeds the criterion (Van Zandt, 2000). These models tend to employ Vickers’ (1979) balance of evidence hypothesis to account for confidence. According to the balance of evidence hypothesis, confidence indexes the difference between the amounts of evidence in the competing accumulators when the response is made. If the difference between the amounts of evidence in the two stores is large, confidence (and likely accuracy) will be high; if the difference is small, confidence (and likely accuracy) will be low.

Again, this logic can be extended to predict a meaningful CA relationship when a witness identifies a lineup member. Once more let us assume that an initial inspection of the lineup reveals one lineup member that the witness believes might be the perpetrator. The witness compares this individual to their memory of the culprit, and this comparison generates some evidence suggesting that the individual is the culprit (i.e., some degree of match between the individual and the witness’ memory of the culprit) and some evidence that the individual is not the culprit (i.e., some degree of mismatch between the individual and the witness’ memory of the culprit). As the degree of match between the lineup member and the witness’ memory of the offender increases, so does the discrepancy between the accumulated evidence favouring the two response alternatives (i.e., this lineup member is the culprit versus this lineup member is not the culprit). As this discrepancy increases, so does the likely accuracy of the decision and the witness’ confidence in that decision.

Although the conceptualisation of confidence processing offered by post-decisional locus models differs from that proposed by SDT–based models, both classes of models suggest that confidence indexes stimulus discriminability. Thus, both classes of models hold that confidence is derived from the same evidence that underlies the decision-making process and, consequently, that conditions facilitating improved memory and accurate responding (e.g., long exposure durations, focused attention, short retention intervals) should also
produce higher confidence. Conversely, conditions that hinder accurate responding should produce lower levels of confidence.

**Research on the confidence-accuracy relationship for eyewitness identification decisions**

Despite sound theoretical support for a positive CA relationship, research initially converged on the view that eyewitness confidence is a poor predictor of identification accuracy. Meta-analyses of correlational investigations of the CA relationship typically reported average correlation coefficients ranging from around zero to .3 or .4 (e.g., Bothwell, Deffenbacher & Brigham, 1987; Cutler, Penrod & Martens, 1987; Sporer, Penrod, Read & Cutler, 1995). Although stronger correlations were reported for participants who chose someone from the lineup (cf. those who reject the lineup; Sporer et al., 1995), the correlational research suggested a moderate relationship at best.

However, there are good reasons to believe that this research inadequately assessed the relationship between confidence and accuracy. These arguments are covered in detail elsewhere (e.g., Brewer et al., 2005; Brewer & Wells, 2006; Juslin, Olsson & Winman, 1996) so are only briefly reviewed here. Researchers have argued that the point-biserial correlation may be an inappropriate statistical index of the relationship (Juslin et al., 1996). When assessing the CA relationship, the point-biserial correlation compares a continuous variable (e.g., confidence as typically assessed on 0%, 10%, 20%, ... 100% scale) with a binary outcome variable (accuracy). Correlation therefore indexes the distance between the distributions of confidence ratings for correct and incorrect decisions (Brewer et al., 2005), comparing confidence to an unrealistic standard of discrimination (Juslin et al., 1996). Others have argued that the typical homogeneity of encoding and testing conditions across participants in eyewitness identification research may artificially constrain variation in memory quality (and, consequently, in confidence and accuracy) undermining CA correlations. These researchers have demonstrated that conditions producing variations in
memory quality produce stronger CA correlations (Lindsay, Nilsen & Read, 2000; Lindsay, Read & Sharma, 1998). Further, research using an alternative method of analysis – calibration – has repeatedly demonstrated (a) positive and generally monotonic relationships between eyewitness confidence and identification accuracy when witnesses identify a lineup member as the culprit and (b) that these relationships often exist alongside typically weak CA correlations (Brewer, Keast & Rishworth, 2002; Brewer & Wells, 2006; Juslin et al., 1996; Olsson, 2000; Olsson, Juslin & Winman, 1998; Palmer, Brewer, Weber & Nagesh, 2013; Sauer, Brewer, Zweck & Weber, 2010; Sauerland & Sporer, 2009). Thus, low to moderate CA correlations may exist in association with meaningful CA relationships, and a consideration of the CA relationship should not rely on correlational analyses.

**CA calibration**

The calibration approach involves plotting the proportion of accurate decisions at each level of confidence (see Figure 1 for a schematic representation). In a typical eyewitness identification experiment, a witness will be shown a crime stimulus and then, after a delay, asked to view a lineup and make an identification decision (i.e., either identify a lineup member as the culprit or reject the lineup as a whole). After making this decision, the eyewitness will be asked to rate their confidence (typically on a 0%-100% scale) in the accuracy of the decision. Researchers can then, across participants, determine the accuracy rates of decisions made at each level of confidence. Confidence and accuracy are perfectly calibrated when, as indicated by the dotted line in Figure 1, 100% of decisions made with 100% confidence are correct, 80% of decisions made with 80% confidence are correct, 50% of decisions made with 50% confidence are correct, and so on. The calibration approach provides useful information about the CA relationship in a number of ways. For example, by plotting variations in accuracy as a function of confidence, a visual inspection of the calibration curve provides an indication of the linearity of the relationship. Also, by
comparing the obtained curve to the ideal, the “realism” of confidence judgements can be assessed for over- or under-confidence. As indicated in Figure 1, functions that fall below the ideal indicate overconfidence, while those above the ideal indicate underconfidence.

Calibration research is common in a variety of human judgment and decision-making domains, but the approach has been used much more sparingly when examining the relationship between eyewitness confidence and identification accuracy. Nevertheless, there is now sufficient research using the calibration approach to support a number of general conclusions about the CA relationship for eyewitness identification. First, when a witness identifies a lineup member as the culprit, there is a generally linear, positive relationship between confidence and accuracy. Second, calibration curves tend to display overconfidence in the upper half of the confidence scale. That is, although accuracy increases systematically with confidence, mean accuracy at each level of confidence tends to be lower than the level of confidence expressed. The degree of overconfidence varies as a function of experimental conditions related to task difficulty (e.g., delay between viewing the crime and viewing the lineup; Sauer et al., 2010), target-absent base rate (Brewer & Wells, 2006), participants’ metacognitive appraisals of their own memory ability (e.g., Brewer et al., 2002) and age (Keast, Brewer & Wells, 2007).

The extant CA calibration literature is generally encouraging regarding the predictive value of eyewitness confidence when a witness identifies a lineup member as the culprit (especially when confidence is very high). However, research also clearly demonstrates a number of robust phenomena that limit this conclusion. For example, even very high levels of confidence do not guarantee that an identification is correct (Brewer & Wells, 2009). For young children calibration is poor, although we do not yet have a firm explanation for this (Keast et al., 2007). And, the systematic CA relationship described above holds for witnesses
who identified a lineup member as the culprit, but not for witnesses who rejected the lineup (see Brewer, 2006; Brewer et al., 2005).

**CA calibration for lineup rejections**

Lineup rejections are common in both lab and field settings (e.g., Behrmann & Davey, 2001; Pike, Brace & Kynan, 2002; Valentine, Pickering & Darling, 2003), and can provide useful information about the likely guilt of a suspect (Clark, Howell & Davey, 2008; Wells & Lindsay, 1980; Wells & Olson, 2002). Consequently, we will devote some discussion here to the perplexing, and under-researched, issue of why confidence is meaningfully related to accuracy for identifications but not for lineup rejections. Again, the theoretical mechanisms underlying confidence processing are informative. Confidence and accuracy should be meaningfully related when they share an evidential basis; when both are based on the degree of match between a presented stimulus (i.e., a lineup member) and the witness’ memorial image of a previously viewed stimulus (i.e., the culprit). Weber and Brewer (2004, 2006) proposed the *information asymmetry hypothesis* to account for the superior CA calibration observed for positive (i.e., I have seen this face before) compared to negative (i.e., I have not seen this face before) face recognition decisions, arguing that the comparison process underlying positive and negative decisions differs. Specifically, a positive decision allows confidence to be based on an assessment of the degree of match between a single test face and a specific face in memory. In contrast, a negative decision indicates that the test face does not match any relevant face in memory. Thus, a negative decision does not provide a one-to-one comparison upon which confidence can be based.

This reasoning may also help explain the consistent discrepancy between CA relations observed for identifications and rejections from lineups. When a witness identifies a lineup member, a comparison of this individual with the witness’ memory of the perpetrator provides a common basis for both the decision and confidence. Confidence and accuracy are
related because both index the quality of the witness’ memory and the degree of match between an individual lineup member and the witness’ memory of the culprit. In contrast, confidence ratings for lineup rejections refer to the rejection of the lineup as a whole, rather than the rejection of individual lineup members, and the basis for rejection confidence is not obvious.

Empirical support for the information asymmetry hypothesis has been mixed. While Weber and Brewer’s (2006) initial investigation using an old-new face recognition paradigm provided little support, recent research using an eyewitness identification paradigm has produced results consistent with the hypothesis. Sauerland, Sagana, and Sporer (2012) used “showups” (a type of identification procedure in which a witness is presented with a single suspect and asked to either identify or reject the stimulus) to ensure that, regardless of whether the witness identified or rejected the suspect, the decision and confidence would be based on a comparison of the degree of match between a single test stimulus and the witness’ memory of the culprit. These conditions produced a monotonic positive relationship between confidence and accuracy for rejections. Recent data from two projects in which one of the authors (JS) has been involved provide additional tentative support. Using face recognition and identification paradigms, and various confidence protocols that required participants to provide confidence ratings for individual rejected faces (cf. rejected lineups as wholes), this research suggests that rejection confidence can be meaningfully related to accuracy when confidence indexes a comparison of the degree of match between a single lineup member with the memorial image of a previously seen stimulus (e.g., Lindsay et al., 2013).

The available evidence is certainly not sufficient to support a general conclusion that confidence is meaningfully related to rejection accuracy, even when rejection confidence is based on comparisons of single test stimuli with single items in memory. However, these early findings highlight two issues for future research. First, alternative protocols for
collecting confidence may provide information that can assist in evaluating the accuracy of lineup rejections. If alternative methods of collecting confidence can increase the diagnostic value of confidence for lineup rejections, then the exculpatory value of lineup rejections will also be improved, and the risks of wrongful prosecution and the mismanagement of investigative resources may be reduced. Second, this research more generally demonstrates that protocols for collecting confidence ratings should be informed by an understanding of the theoretical basis for the CA relationship. Specifically, when protocols for collecting confidence are designed to permit an assessment of the degree of match between lineup members and the witness’ memory for the culprit, the predictive value of confidence is likely to increase. The importance of designing confidence protocols that allow confidence to reflect an assessment of match becomes even clearer when we consider non-memorial factors that can influence confidence ratings, and weaken the CA association.

**The malleability of confidence**

Confidence is vulnerable to a variety of non-memorial (e.g., social and environmental) influences that reduce the correspondence between the evidential bases for confidence and accuracy, and undermine the CA relationship. For example, post-identification feedback is a powerful source of non-memorial influence, with research consistently demonstrating that post-identification feedback affects confidence for recognition decisions and undermines the CA relationship (Bradfield, Wells & Olson, 2002; Luus & Wells, 1994; Semmler & Brewer, 2006; Semmler, Brewer & Wells, 2004; Wells & Bradfield, 1998, 1999). After making an identification decision, a witness may receive feedback from various sources. The lineup administrator may indicate that the witness has picked the suspect (or someone other than the suspect), and this may be communicated through explicit, verbal means or inferred by the witness from the administrator’s non-verbal behaviour (e.g., body language or facial expression). Alternatively, following discussions
with a co-witness, a witness may learn that the co-witness provided the same (or a different) response. Confirming feedback generally inflates witness confidence, while disconfirming feedback generally reduces confidence. Importantly, these feedback effects on confidence do not depend on the witness’ decision type (i.e., identification or rejection), the nature of the feedback (e.g., confirmation that the witness selected the suspect or simply confirmation that another witness made the same decision), the source of the feedback (e.g., administrator, co-witness, or even computer-based feedback), or whether the feedback is verbal or non-verbal (Garrioch & Brimacombe, 2001).

Research investigating feedback effects on confidence has demonstrated that these effects are robust. However, research also suggests that these effects are likely to be strongest when the witness’ memory of the event is weak, and the relative paucity of internal cues for confidence may increase witness reliance on external cues (Bradfield et al., 2002; Wells & Bradfield, 1998, 1999). For example, conditions that encourage accurate responding tend to reduce the influence of feedback on confidence (Brewer et al., 2005). Further, feedback effects tend to be less pronounced when the witness is able to make their decision quickly (Semmler & Brewer, 2006; Semmler et al., 2004). Presumably this is because quick decisions tend to indicate a strong memorial basis for the decision (i.e., a clear match between a lineup member and the witness’ memory of the offender, or the clear absence of any match between lineup members and the witness’ memory). Nonetheless, the fluency of a quick recognition decision and the quality of the underlying memorial information may provide salient internal cues for confidence, reducing reliance on external cues. Witnesses’ expectations about the likely accuracy of their decisions also appear to moderate feedback effects on confidence. Confidence inflation effects tend to be reduced when the feedback agrees with a witness’ prior expectations regarding the accuracy of their decision (Semmler & Brewer, 2006). Again this suggests that when internal cues to the accuracy of the decision are accessible, external
cues to confidence are less influential. The effects of feedback on confidence can also be reduced by explicitly increasing the accessibility or salience of internal cues to confidence (e.g., by having witnesses think about their confidence before feedback is provided, Wells & Bradfield, 1999) or by undermining the perceived credibility of the feedback source (e.g., by creating suspicion that the feedback provider may have attempted to mislead the witness, Neuschatz et al., 2007; or that the feedback provider was mistaken, Quinlivan, Wells & Neuschatz, 2010). Together, these results converge to suggest that witnesses will attempt to draw inferences from both internal and external cues when assessing confidence, especially when (a) internal cues are weak or inaccessible and (b) external cues are seen as valid sources of information. However, if confidence is to offer useful information about the reliability of an identification decision, it must index the witness’ memory. Thus, the influence of external cues must be minimised.

The established effects of post-identification feedback on confidence have clear implications for practice. First, as recommended by the National Institute of Justice Guidelines for the Collection of Eyewitness Evidence (Technical Working Group for Eyewitness Evidence, 1999), confidence should be assessed immediately following the decision, and before a witness interacts with lineup administrators, investigating officers, or other witnesses. This objective is, of course, most easily realised if lineup administration and recording of witness responses occurs on a computer. Our concern in this chapter is not with the actual method of lineup administration; however, arguments have been made elsewhere for computerised lineup administration and the advantages this has for recording confidence judgments (see, for example, Brewer, 2011). Confidence is an influential index of witness credibility and identification reliability. Thus, feedback effects that inflate confidence (but, by their nature, cannot improve accuracy) distort the CA relationship and can make a witness and their identification appear more compelling than they should. Second, the confidence
rating recorded immediately after the decision is made should be the ‘confidence evidence’
tendered in court. Unsurprisingly, preparing witnesses for cross-examination, and having
witnesses answer the same question repeatedly, increases confidence (Shaw, 1996; Shaw &
McClure, 1996; Wells, Ferguson & Lindsay, 1981). Thus, in-court expressions of confidence
are likely to bias evaluations of identification credibility (see case study). One approach to
combat the confidence-inflating effects of post-identification feedback and pre-trial
preparation would be to present confidence as measured at the time of the identification
alongside any in-court assessment of identification confidence. When their attention is drawn
to confidence inflation, jurors are able to adjust their perceptions of witness and identification
credibility (Bradfield & McQuiston, 2004). However, research also demonstrates that jurors
do not appropriately adjust perceptions of witness credibility in light of apparent confidence
inflation when the witness offers a plausible justification for their increased confidence (e.g.,
by attributing their lower initial confidence to nervousness at the time of the identification,
and/or explaining that they have since recalled additional, confirming information; Jones,
Williams & Brewer, 2008). Thus, from the perspective of preserving the informational value
of confidence ratings it would be reasonable to argue that only confidence recorded
immediately following the decision should be tendered as evidence in court, even though any
such recommendation would likely attract considerable opposition from within the legal
system.

In sum, the literature supports the following conclusions about the CA relationship for
eyewitness identifications. Eyewitness confidence and identification accuracy are
meaningfully related when (a) an adult witness positively identifies a lineup member as the
culprit and (b) confidence is assessed immediately after the decision is made, and prior to any
interaction with the lineup administrator or other witnesses (Brewer, 2006; Brewer et al.,
2005; Brewer & Wells, 2009).
Applied factors that limit the effective use of confidence

A major factor limiting the effectiveness of confidence as an index of accuracy for identification decisions may be the absence of established protocols for collecting witness confidence ratings in a way that maximises their informativeness. The National Institute of Justice Guidelines for the Collection of Eyewitness Evidence (US) recommend that records of the witness’ decision include “the witness’ own words regarding how sure he/she is” (Technical Working Group for Eyewitness Evidence, 1999, p.38). Similarly, in other Western legal systems (e.g., the UK, and Canada) administrators are told to record anything the witness says regarding their identification (Police and Criminal Evidence Act Code D, 2011; Sherrin, 2007). Thus, if the witness volunteers an assessment of confidence it will be recorded, but administrators are not instructed to probe for confidence. This means that there is no guarantee that confidence will be assessed. Moreover, the way in which any obtained confidence rating is expressed may limit its usefulness. We discuss each of these issues in turn.

Why is it necessary to explicitly assess witness confidence? The obvious answer is identifications are prone to error, and confidence may help discriminate between identifications that are more or less likely to be reliable. However, in a broader sense, confidence is valuable because (when measured appropriately) it provides additional information about the quality of the witness’ memory and the extent to which the identified lineup member matches the witness’ memory of the culprit. A positive identification of the suspect may be extremely attractive to police and prosecutors, but in no way guarantees the suspect is the culprit. It is reasonable to assume that a suspect identification indicates that, of the presented lineup members, the suspect provides the best match with a witness’ memory of the culprit. However, procedural and social factors can affect a witness’ decision criterion, making them more or less likely to pick someone from the lineup, without affecting the
degree of match between individual lineup members and the witness’ memorial image of the culprit. Thus, although a suspect identification probably indicates that the suspect provides the best available match, it is much less informative about the quality of this match in an absolute sense. Confidence can provide additional information here. Further, research in laboratory and field settings, together with inferences drawn from the limited available evidence concerning confidence in identifications in actual criminal cases, clearly demonstrates that identifications are made with varying degrees of confidence (e.g., Brewer & Wells, 2006; Sauer et al., 2010). Given the theoretical relationship between confidence and memory quality, variations in confidence can be viewed as indicating qualitatively different degrees of recognition. Thus, it would be foolish not to consider confidence (i.e., to effectively treat all identifications as equally certain) when evaluating identification evidence. Confidence should be viewed as a further source of information about the witness’ memory, rather than simply as a potential index of accuracy.

However, if confidence is to provide this useful information, methods for assessing confidence must (a) attenuate non-memorial influences on confidence, thereby allowing it to index the witness’ memory quality and (b) record confidence in a way that permits meaningful interpretation. How, then, should witness confidence be measured? As already indicated, confidence must be assessed immediately following the identification. Preferably, this process would be run blind for the lineup administrator (and possibly automated) to avoid the influence of any explicit or implicit verbal or non-verbal feedback.

A second issue to be considered is the type of scale used to measure confidence. When guidelines discuss recording confidence ratings, they typically state or imply that confidence should be recorded in the witness’ own words. However, such assessments are likely to be of limited value as spontaneous utterances are likely to be ambiguous. Consider a witness who says “I’m not certain, but I think it is number four”. This statement could apply
equally to a witness who would have responded with 90% or 40% given a numerical confidence scale and, thus, this expression of confidence offers little discrimination regarding the quality of the witness’ memory, or the likely reliability of the identification. Furthermore, in conversation, meaning is added to such utterances through tone of voice, and accompanying non-verbal behaviours. This contextual information may be difficult to preserve in a written record, further reducing the informativeness of the confidence rating. The motivation for current practice - assessing confidence in the witness’ own words - may share a foundation with concerns expressed by researchers over a witness’ ability to translate internal confidence states onto numerical scales (Soll & Klayman, 2004, Windschitl & Wells, 1996). Researchers have argued that although conceptualising certainty numerically may be natural for statistically-trained individuals, this may not be true for the general population, and that verbal scales (e.g., using anchor points such as likely, very likely, extremely likely, certain) may better approximate internal confidence states. Despite these concerns, and encouraged by the similar CA calibration patterns reported for identification decisions regardless of whether numerical and verbal scales were used (Weber, Brewer & Margitich, 2008), we recommend assessing confidence on a standardised, numerical scale (e.g., 0-100%, 1-10, etc.). There is sufficient evidence to demonstrate that adult participants can effectively use numerical scales to discriminate between decisions for which they have higher and lower confidence, and decisions that are more or less likely to be accurate (e.g., Brewer & Wells, 2006; Juslin et al., 1996; Sauer et al., 2010; Sauerland & Sporer, 2009; Weber & Brewer, 2003, 2004). An interval scale also offers an internal hierarchy of responses and provides a less ambiguous metric for those interpreting the reported level of confidence. In contrast, a confidence response such as “rather likely” or “quite likely” (which may come from a witness’ own words or from responses on a verbal confidence scale) is more difficult to interpret. For example, these responses could conceivably indicate levels of confidence
falling toward the middle or the upper end of the confidence continuum. Admittedly, fine-grained numerical confidence scales may not be appropriate for all eyewitnesses (e.g., young children, people with learning difficulties). Keast et al. (2007) found that children in the 11-13 age range appeared just as capable as adults at interpreting such confidence scales although, despite that, their confidence judgments were not diagnostic of identification accuracy (Keast et al. 2007). However, such scales are unlikely to be suitable for younger children.

When research clearly indicates that (a) confidence is highly influential and (b) confidence can be informative under certain conditions, and some courts even require that confidence be recorded (e.g., State v. Henderson, 2011) and considered when evaluating identification evidence (e.g., Neil v. Biggers, 1972), why would criminal justice systems be reluctant to establish systematic protocols for eliciting, recording and maximising the informational value of witness confidence ratings? We can only speculate here. Perhaps practitioners believe, contrary to the scientific literature, that recognition memory is an “all or nothing” process. The witness both recognises (and identifies) the suspect or they do not, and confidence ratings are irrelevant because they offer no additional information regarding the validity of recognition. Or perhaps practitioners may believe, again contrary to empirical evidence, that a witness would not identify a lineup member if they were uncertain. Alternatively, if practitioners acknowledge that some identifications are made with less than complete certainty, perhaps they believe that uncertain identifications are of no value. Specifically, prosecutors may feel that an identification made with, for example, 70% confidence will be easily undermined by the defence. They may be right. However, it is certainly true that to not consider confidence – and instead present all identifications as implying certainty, despite all evidence to the contrary – undermines the informational value of the evidence. If such beliefs contribute to the desire to avoid collecting explicit confidence
ratings, they systematically undermine efforts to appropriately evaluate identification evidence and increase the risk of miscarriages of justice.

If legal decision-makers are concerned that assessing confidence may undermine prosecutorial efforts, the answer may be to correct faulty perceptions of identification evidence within the legal system rather than avoid the issue and risk misrepresenting the diagnostic value of identification evidence. As Brewer and Wells (2009) note, rather than treating an identification as a clear-cut indication of guilt, it is more useful to view the identification as an indication that a suspect is worth further investigation or, if there is sufficient corroborating evidence, that the suspect should be charged. Similarly, in court, although jurors view identification evidence (particularly when provided by confident witnesses) as a compelling indication of defendant guilt, they should be encouraged to view the identification as one piece of probabilistic evidence that must be considered alongside the other evidence when assessing the defendant’s guilt. During investigations and at trial, confidence can enrich this source of information to help guide the decision-making process regarding the likely guilt of the accused.

A novel use for eyewitness confidence: Confidence as an index of recognition

In line with the re-conceptualisation of identification evidence suggested above, we now discuss a radical departure from traditional identification practice. Although this approach involves a drastic change in the way identification evidence is collected and interpreted, we argue that it provides a more informative index of recognition, and represents a more valid approach to testing witness memory. As discussed previously, various theories of confidence processing for recognition memory hold that confidence indexes the degree of match between a presented item and an image in memory. This suggests an interesting possibility for assessing witness recognition. Specifically, avoiding explicit, categorical identification responses (and thereby attenuating non-memorial influences on witnesses’
decision criteria) and instead simply asking the witness to rate their confidence (on a scale from 0-100%) that each lineup member is the culprit (henceforth, *ecphoric confidence ratings*) may provide a more sensitive and informative index of recognition, and a more direct assessment of the degree of match between individual lineup members and the witness’ memory of the culprit (Brewer, Weber, Wootton & Lindsay, 2012; Sauer, Brewer & Weber, 2008; Sauer, Brewer & Weber, 2012; Sauer, Weber & Brewer, 2012).

Early investigations of the diagnostic value of ecphoric confidence ratings have provided a number of encouraging findings. First, basic face recognition research revealed a generally linear, positive relationship between ecphoric confidence ratings and the likelihood that a face has been previously seen (Sauer et al., 2012b). Second, after the application of classification algorithms to determine when a confidence rating or pattern of confidence ratings could be treated as a positive identification of the suspect (described in detail in Sauer et al., 2008), ecphoric confidence ratings consistently produced higher classification accuracy than categorical recognition decisions in face recognition tasks, and from simultaneous and sequential lineups (Brewer et al., 2012; Sauer et al., 2008; Sauer et al., 2012a; Sauer et al., 2012b). That is, confidence ratings were more effective than participants’ decisions at indicating whether or not participants had seen a presented face or lineup member before.

Finally, analyses of classification performance using SDT-based measures of discrimination (i.e., the sensitivity of the index of recognition) and response bias (i.e., the general tendency to return an increased or decreased number of positive classifications) demonstrated that the improved classification accuracy associated with the use of ecphoric confidence ratings (cf. categorical decisions) was attributable to improvements in discrimination rather than any change in bias (Sauer et al., 2012b). Compared to participants’ recognition decisions, ecphoric confidence ratings provided a more sensitive index of recognition.
Collapsing confidence ratings into categorical classifications in order to compare classification performance using ecphoric ratings with performance based on categorical recognition decisions permits a valuable demonstration of the diagnostic value of the confidence procedure. However, it also reduces the richness of the recognition information provided. Brewer et al.’s (2012) analyses included an additional and more informative treatment of the data. Specifically, for each lineup, the researchers determined whether there was a single highest, maximum confidence value. If there was a maximum confidence value, the researchers examined variations in the likely guilt of the suspect as a function of the discrepancy between the maximum and next-highest values. Across three experiments, this profile analysis produced striking results. First, consistent with models that conceptualise the strength of, or confidence in, recognition as an index of the relative similarity of decision alternatives to a previously viewed item (e.g., Valentine, 1991), the likely guilt of the suspect increased almost monotonically as a function of the discrepancy between the maximum and next-higher confidence ratings. Second, when this discrepancy was large (e.g., ≥ 80%) the likely guilt of the suspect was very high (e.g., 80-100%) and, until the discrepancy fell to 30-50%, confidence ratings were a better predictor of suspect guilt than were categorical identification decisions.

These findings have two important implications. First, as Brewer et al. (2012) note, they extend demonstrations of group-level differences in classification accuracy to show that patterns of confidence ratings can offer reliable diagnostic information about suspect guilt for individual witnesses. Second, the monotonic positive relationship between the discrepancy measure and the likely guilt of the suspect hints at the plausibility of avoiding categorical classifications entirely, in favour of a probabilistic treatment of identification evidence. Traditional identification testing approaches basically assign a witness’ recognition memory ‘output’ to categories such as “It’s number 6” or “He’s not there” or “I’m not sure.” The legal
system may benefit, however, from maximising the amount of information available from the witness’ memory, and considering what this information says about the likely guilt of the suspect/defendant.

Although the boundary conditions for the confidence procedure described above clearly require further investigation, this procedure has the potential to address a number of the problems that have plagued traditional identification practices. First, avoiding explicit, categorical decisions attenuates the various non-memorial influences on criterion placement that compromise the extent to which the eventual decision reflects the degree of match between a lineup member, or members, and the witness’ memory of the culprit (i.e., the construct of interest), and contribute to mistaken identifications. Second, this approach provides legal decision-makers with a richer source of information upon which to base assessments of likely guilt. As mentioned previously, a single decision provides little information about the extent to which the identified lineup member matches the witness’ memory of the culprit. Further, this decision provides no information about the extent to which the identified lineup member is favoured over the alternatives and, if the suspect is not identified, the eventual decision provides no information about the degree of match between the suspect and the witness’ memory of the culprit (other than that the degree of match did not exceed the criterion for identification in the case of a lineup rejection, or that the suspect was not the best match in the case of a foil identification). In contrast, in all cases, the confidence procedure provides investigators with useful information about (a) the extent to which the suspect matches the witness’ memory of the culprit and (b) the similarity of the suspect to the witness’ memory, relative to other lineup members.

As Brewer and Wells (2011) recently argued, although a procedure that does not involve a witness actually picking or rejecting a suspect is likely to encounter strong resistance from the police and courts, “... any procedure that reduces the likelihood that
culprits go free and innocent people are convicted warrants serious attention from a research perspective and from the perspective of giving away psychological science” (p. 26). We believe that some variant of the confidence procedure outlined above may well offer an opportunity to abandon the problematic conceptualisation of identification evidence as some absolute indication of guilt, in favour of a more scientific consideration of identification evidence as yet another piece of probabilistic evidence – evidence which must be considered alongside the available corroborating evidence when evaluating the likely guilt of a suspect/defendant.

References


Footnotes

1Note that recent judgments (e.g., State v Henderson, 2011; Oregon v. Lawson, 2012) discuss the CA relation, but the emphasis of their recommendations is primarily on recording confidence immediately after the decision to combat the inflationary effects of feedback on confidence, not because confidence recorded at that point can provide a pointer to the reliability of the identification.

2Note that it is possible that the presence of two or more lineup members who closely match the witness’ memory of the offender might reduce confidence in the accuracy of the eventual decision, despite a high degree of match between the selected lineup member and the witness’ memory of the offender. Ongoing research in our laboratory addresses this issue but does not yet permit reliable conclusions.
Figure 1. Schematic representation of confidence-accuracy (CA) calibration displaying perfect calibration, overconfidence and underconfidence.