Chapter 10

Eyewitness Identification

James D. Sauer & Matthew A. Palmer
University of Tasmania
Neil Brewer
Flinders University

Supported by funding from Australian Research Council grants DP150101905 to Neil Brewer et al. and DP140103746 to M. Palmer et al.

To be cited as:

Sauer, J.D., Palmer, M.A., & Brewer, N. (2018, in press). Eyewitness identification. In N. Brewer & A. Douglas (Eds.), Psychological Science and the Law. Guilford.

After viewing a crime (or other event of interest) an eyewitness will often be presented with some form of identification task (either live or photo array), and asked whether they recognize someone from the lineup as the person of interest from the initial event. The lineup will generally include a suspect (who may or may not be the culprit) and a number of fillers (individuals known to be innocent). The witness can either identify the suspect, identify a filler, reject the lineup (i.e., decline to identify anyone) or, in some cases, indicate that they are unable to make a decision (i.e., respond that they “don’t know”). The witness’s response can have important consequences for the ongoing investigation and, more broadly, for attempts to prosecute the guilty. If the witness identifies the suspect, the likelihood of the suspect being prosecuted increases. If the witness rejects the lineup, the police may decide to redirect their investigative efforts to pursue an alternative line of enquiry or look for an alternative witness. Eyewitness identification evidence is both compelling and prone to error (Steblay, Dysart, Fulero, & Lindsay, 2003; Steblay, Dysart, & Wells, 2011). Given the weight placed by triers of fact on identification evidence, it is unsurprising that false identifications (i.e., of innocent suspects) are a leading cause of wrongful conviction in many jurisdictions (cf. Innocence Project, 2017). Moreover, a failure to identify the culprit, if present in the lineup, can undermine investigative and prosecutorial efforts. An awareness of these consequences has motivated a substantial body of research literature aimed at improving our understanding of the causes of identification error and evaluating various imaginative attempts to mitigate these errors.

In this chapter we summarize what we consider are the major findings to emerge from the now considerable literature. In many cases, however, there already exist substantial reviews or meta-analyses and, consequently, we only review these areas quite briefly, noting the main findings and pointing readers in the direction of major reviews. This applies especially to the consideration of variables that are known to affect identification performance and yet are outside the control of justice system professionals charged with administering lineups. In contrast, we devote more attention to a number of important questions for which, in many cases, we cannot provide conclusive answers based on the current state of the literature. Our objective here is to prompt a critical re-consideration of what is known, what is unknown, and how we might best advance the use of psychological science to benefit practitioners in the criminal justice system.

Things We Know About Identification Test Performance but Cannot Change

The research literature identifies a number of factors important to understanding identification performance, although it is important to bear in mind that many of them (often referred to as estimator variables, Wells, 1978) are outside the control of the justice system. An identification is a recognition memory task. Thus, factors at encoding, or between the encoding and test phases, that affect memory quality tend to show predictable effects on identification accuracy. For example, increased exposure durations and better viewing conditions (e.g., shorter viewing distances) tend to be associated with improved recognition performance (e.g., Lindsay, Semmler, Weber, Brewer, & Lindsay, 2008; Memon, Hope, & Bull, 2003; Palmer, Brewer, Weber, & Nagesh, 2013). Similarly, divided (cf. full) attention at encoding – whether prompted by the presence of a weapon (Steblay, 1992) or some more general mechanism (e.g., Palmer et al., 2013) – is also associated with reduced identification performance. There is also evidence that, consistent with basic memory tasks, stimulus distinctiveness is associated with improved face recognition, though much of this evidence comes from basic face recognition tasks, rather than eyewitness identification tasks (e.g., Light, Kayra-Stuart, & Hollander, 1979; Sauer, Brewer, & Weber, 2008; Semmler & Brewer, 2006). There is also a general tendency for people to be better able to identify faces of their own (cf. another) race (the *cross-race effect*; Meissner & Brigham, 2001), and for longer retention intervals between the crime and the identification test to be associated with poorer identification performance (e.g., Palmer et al., 2013; Sauer, Brewer, Zweck, & Weber, 2010). Witness characteristics also show reliable effects on identification performance, with the most striking example being the tendency for child witnesses to be less accurate than adults (Fitzgerald & Price, 2015). The effect of all of these variables is manifested in either a lower chance of a correct identification when the culprit is present in the lineup, or a higher likelihood of an erroneous identification decision (i.e., an innocent suspect or filler pick) when the culprit is not present, or both of these outcomes.

These effects are all intuitive and well-grounded in memory theory. Moreover, an appreciation of their nature is important from the perspectives of understanding identification decision making and evaluating the likely reliability of identification evidence. But we must add several caveats. First, knowing that identification performance varies in a predictable manner with changes on these variables does not mean that the accuracy of any individual identification test outcome can be "diagnosed". For example, knowing that identification performance deteriorates as the retention interval between crime and identification test lengthens does not allow the conclusion that a particular identification made after a particular interval (e.g., 3 days or 3 months) will be accurate or inaccurate. Or, knowing that child witnesses are more likely to choose from a culprit-absent lineup does not mean that, if a child witness picked the police suspect from the lineup but an adult witness didn't, the police suspect must be innocent. Second, in a number of cases, the generality of these effects across stimulus materials has not been established. Thus, it is unclear how dependent these effects are on the idiosyncratic properties of the stimuli and testing protocols for which they have been observed. Third, even in cases where "main effects" are robust, the literature provides a limited understanding of the boundary conditions for these effects, or the extent to which these effects might be moderated by other factors of applied and theoretical relevance. For example, increased exposure duration might attenuate deleterious effects on identification performance related to the distracting presence of a weapon at encoding or a very long retention interval. We explore these caveats in more depth in the section on generalizing findings from the lab environment to applied settings.

Predicting Identification Accuracy

Given that identification errors are common, researchers have attempted to identify independent markers of identification accuracy. Although a variety of approaches to indexing identification accuracy have been pursued (e.g., phenomenological reports, Dunning & Stern, 1994; Palmer, Brewer, McKinnon, & Weber, 2010; eye movement patterns, Mansour & Flowe, 2010; Mansour, Lindsay, Brewer, & Munhall, 2009), we focus on the two most-studied markers of accuracy: eyewitness confidence and response latency (i.e., the time taken to make the identification response). Below we consider the utility of these factors as markers of accuracy of identification decisions.

Confidence and Accuracy for Eyewitness Identifications

Eyewitness confidence exerts a powerful influence on decision-making in legal settings. Police, lawyers, and jurors believe confidence is reliably linked to accuracy (Deffenbacher & Loftus, 1982; Potter & Brewer, 1999). Further, experimental manipulations of witness confidence affect mock-jurors' perceptions of witness credibility and defendant guilt (Bradfield & Wells, 2000; Cutler, Penrod, & Dexter, 1990).

More importantly, there is compelling theoretical support for a positive confidence-accuracy relationship. Various theories of confidence processing – emerging from a variety of human judgment and decision-making domains (see Horry & Brewer, 2016, for a review) – hold that confidence and accuracy share an evidential basis related to memory quality and stimulus discriminability. For example, in a recognition memory task (e.g., a lineup), an individual will typically compare a presented test stimulus (e.g., a lineup member) with a memorial image of a previously-viewed stimulus (e.g., a culprit). This comparison generates some degree of evidence that the two stimuli match. This evidence forms the primary basis for both the decision and confidence, and this shared evidential basis supports a positive confidence-accuracy relationship. As the quality of the witness's memory and the degree of match between an identified lineup member and the witness's memory of the culprit increase, so do the likely accuracy of and the witness's confidence in that decision.

Despite strong theoretical support for a positive confidence-accuracy relation, meta-analyses of correlational investigations of the confidence-accuracy relationship suggested a moderate relationship at best (reporting average coefficients between zero and .4, e.g., Sporer, Penrod, Read, & Cutler, 1995). These findings may have motivated the scepticism about the confidence-accuracy relationship among eyewitness researchers (e.g., 73% of surveyed experts being willing to testify that confidence is not a reliable predictor of identification accuracy; Kassin, Tubb, Hosch, & Memon, 2001). However, researchers have subsequently argued that the point-biserial correlation is an inappropriate index of the confidence-accuracy relation (e.g., Juslin, Olsson, & Winman, 1996), and demonstrated repeatedly (using an alternative method of analysis: *calibration*) that robust confidence-accuracy relations often co-exist with typically weak confidence-accuracy correlations (e.g., Brewer & Wells, 2006; Palmer, Brewer, Weber, & Nagesh, 2013; Sauer, Brewer, Zweck, & Weber, 2010). The calibration approach involves plotting the proportion of accurate decisions for each level of confidence. Perfect confidence-accuracy calibration is obtained when 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. Visual comparison of the obtained and ideal calibration functions (together with associated statistical indices) provides information about the linearity of the relationship, and tendencies toward over- or under-confidence (for further detail, see Brewer & Wells, 2006, or Juslin et al., 1996).

The extant literature on confidence-accuracy calibration demonstrates, for choosers (i.e., witnesses who identify a lineup member as the culprit), a generally linear, positive relationship between confidence and accuracy (Brewer & Wells, 2006 Palmer et al., 2013; Sauer et al., 2010; Wixted, Mickes, Dunn, Clark, & Wells, 2016; Wixted & Wells, 2017). As confidence increases, so does the likely accuracy of the identification. Thus, confidence can provide useful information about the reliability of an identification. However, the literature provides a number of important caveats to this conclusion. First, this relationship typically displays overconfidence. Although accuracy increases systematically with confidence, mean accuracy at each level of confidence tends to be lower than the level of confidence expressed. Further, overconfidence (a) increases as a function of task difficulty (Palmer et al., 2013; Sauer et al., 2010) and target-absent base rates (i.e., the proportion of occasions in which the culprit is not present; Brewer & Wells, 2006), (b) can be large for child witnesses (Keast, Brewer, & Wells, 2007), and (c) is influenced by participants' meta-cognitive beliefs about their memory ability (Brewer, Keast, & Rishworth, 2002). Second, and following from the above, very high levels of confidence do not guarantee accuracy (Brewer & Wells, 2009). Third, the linear confidence-accuracy relation observed for choosers does not hold for non-choosers (i.e., witnesses who reject the lineup). Finally, confirming post-identification feedback can inflate confidence and, in turn, undermine the confidence-accuracy relationship (Semmler, Brewer, & Wells, 2004; Wells & Bradfield, 1998; 1999).

Post-identification feedback can be obtained from a variety of sources (e.g., lineup administrators and co-witnesses), and may be communicated explicitly (e.g., "Good, you identified the suspect") or inferred from non-verbal behaviour (e.g., lineup administrators' facial expressions). Thus, to be informative about the reliability of an identification decision, confidence must be assessed immediately following the decision, and prior to any witness interaction with lineup administrators or co-witnesses. Moreover, to preserve the informational value of confidence ratings, we would argue that only confidence recorded immediately following the decision should be tendered as evidence in court (Sauer & Brewer, 2015). Although such a recommendation would likely attract considerable opposition from within the legal system, it is critical that such a practice becomes commonplace if confidence is to inform assessment of identification reliability. Even so, we note that any procedural factors (i.e., biases) that influence confidence but not accuracy may still undermine the confidence-accuracy relation.

Despite robust empirical support for a meaningful relationship between confidence and accuracy, the absence of established protocols for systematically collecting and preserving witness confidence ratings in most criminal justice systems currently represents a significant practical hurdle to the effective use of confidence as an index of identification accuracy (Sauer & Brewer, 2015). However, this problem could easily be remedied via computerized lineup administration incorporating a built-in request for a confidence judgment following the identification decision (Brewer, 2011).

Response Latency and Accuracy for Eyewitness Identifications

As with confidence, there are strong theoretical grounds for predicting a relationship between response latency and accuracy. A strong (cf. weak) memorial representation of the culprit, and a lineup member who provides a good (cf. poor) match to this memory, should promote recognition (a largely automatic process) and, consequently, faster responding with increased accuracy (e.g., Sporer, 1992; 1993). The extant literature supports these predictions, consistently demonstrating lower response times for accurate (cf. inaccurate) identifications (e.g., Brewer, Caon, Todd, & Weber, 2006; Dunning & Perretta, 2002; Sporer, 1994). However, despite robust evidence for a negative latency-accuracy relationship, two points are worth noting. First, eyewitnesses can operate at any point on the speed-accuracy continuum. Thus, individual differences in decision-making may muddy the latency-accuracy relationship in applied settings. For example, one witness may have a strong recognition experience and respond quickly and accurately, while another may have the same initial recognition experience and settle quickly on their preferred candidate, but spend additional time interrogating this initial preference before offering a (correct) overt response. Alternatively, a witness may be uncertain, but guess quickly and incorrectly. Thus, a slow response does not guarantee an error and a quick response does not guarantee accuracy.

Second, and related to the previous point, the absence of a reliable metric indicating when a response is "quick enough" to indicate accuracy severely limits the applied utility of latency as a marker of identification accuracy.

Some early research suggested that specific latency "windows" might reliably diagnose identification accuracy, at least for simultaneous lineups. For example, Smith, Lindsay, and Pryke (2000) reported an accuracy rate of ≈70% for identifications made in under 16 s, compared to accuracy rates of ≈43% and ≈18% for identifications made in 16 – 30 s and over 30 s, respectively. Dunning and Perretta (2002) then reported that, across multiple experiments, identifications made within a 10 – 12 s time boundary showed very high accuracy rates (≈87%) compared to identifications made outside this boundary (≈50%). However, subsequent research seriously challenged the generalizability of these time boundaries and the associated accuracy rates. First, across a number of large-scale experiments using identical encoding and test stimuli, Weber, Brewer, Wells, Semmler, and Keast (2004) demonstrated that the time boundary that best discriminated correct from incorrect identifications varied considerably (from 5 to 29 s). Further, the accuracy rates for decisions made within and outside optimum time boundaries were much lower than those reported by Dunning and Perretta (with accuracy rates ranging from ≈20-79% before the boundary, and ≈11-56% after the boundary). Brewer et al. (2006) also demonstrated that (a) optimum time boundaries could be experimentally manipulated (via manipulations that affect stimulus discriminability), and (b) accuracy rates associated with optimum time boundaries were again lower than those reported by Dunning & Perretta. Finally, Sauer, Brewer, and Wells (2008) were unable to identify a stable latency-based metric for diagnosing the reliability of identifications made from sequential lineups. Thus, despite sound theoretical and empirical support for a negative latency-accuracy relationship, variability in empirically-derived optimum time boundaries and the diagnostic value of these boundaries undermines the utility of response latency as an index of accuracy in applied settings. Nonetheless, latency may contribute to evaluations of identification evidence if viewed as an index of memory quality rather than simply identification accuracy.

Confidence and Latency Combined as Indices of Memory Quality

Neither confidence nor latency provide a foolproof method for diagnosing identification accuracy. However, both – especially when considered together – can provide useful information about the quality of a witness's memory, the strength of their recognition experience and, consequently, the informational value of the identification evidence. Various theoretical frameworks propose confidence and latency index memory strength and stimulus discriminability (e.g., Vickers, 1979). Thus, provided the lineup is fair, if a witness identifies the suspect quickly and with high confidence, this likely indicates that the witness's memory for the culprit is strong, that the suspect matches this memory well, and that the identification is more likely to indicate suspect guilt. Consistent with this prediction, in lab settings, studies have demonstrated impressive levels of accuracy for rapid identifications made with high confidence (Brewer & Wells, 2006; Sauerland & Sporer, 2009; Weber et al., 2004).

However, in applied settings, we generally cannot establish ground truth (cf. in the absence of supporting DNA evidence). We must infer likely guilt from the identification evidence, rather than assess the identification against a known state of the world (i.e., suspect guilt or innocence). Thus, when discussing methods for evaluating identification, thinking in terms of these methods' ability to diagnose accuracy potentially fosters an overly simplistic way of conceptualizing identification evidence in these settings. Thus, as per Brewer and Weber (2008), we suggest that the value of considering confidence and latency lies not in their ability to definitively diagnose accuracy but, rather, in their potential to add information about the quality of memorial evidence underlying the identification decision. This point leads to two important considerations when thinking about how identification evidence should be collected and interpreted. First, despite the apparent clarity of an identification as an indication of suspect guilt, we must bear in mind that recognition is not an "all or nothing" process. Memory is fallible, and recognition decisions reflect both the quality of memorial evidence available and the individual's decision criterion. This criterion can vary according to social factors unrelated to memory quality. Thus, an identification is not a clear-cut indication of guilt and, consequently, should be interpreted probabilistically and alongside other forms of forensic evidence when assessing the likely guilt of a suspect. An identification is just one piece of evidence against the suspect, and the value of this evidence depends on the quality of the witness's memory (and the quality of procedures used to obtain the identification). As indices of memory quality, confidence and latency can help inform an assessment of the evidentiary value of an identification. Second, an identification decision on its own is less informative than it may appear. Although an identification probably indicates that, of the presented lineup members, the selected person provides the best match to the witness's memory of the culprit, it says nothing about the strength of that match or the extent to which the selected individual was favored over the alternatives (i.e., the witness's ability to discriminate a culprit among fillers). This second point leads to proposed alternative to traditional identification tests, discussed in a later section.

Things We Know and Can Change About Identification Test Performance

Presentation of the Lineup

Some aspects of lineup administration are non-controversial and supported by the vast majority of eyewitness researchers. One example is the use of single suspect lineups, whereby one lineup member is the police suspect and all other lineup members are known-to-be innocent fillers. Compared to multiple-suspect lineups, single suspect lineups reduce the incidence of false identification because they allow incorrect filler identifications to be classified as known errors (Wells & Turtle, 1986; see Wells, Smalarz, & Smith, 2015 for a detailed discussion of *filler siphoning*). In contrast, multiple-suspect lineups—and especially all-suspect lineups (in which all lineup members are suspects)—dramatically increase the chances that an innocent person will be prosecuted.

Double-blind testing—whereby the lineup administrator does not know which lineup member is the police suspect—is another example (for a detailed discussion, see Kovera & Evelo, 2017). Single-blind testing, whereby the administrator but not the witness knows which lineup member is the suspect, leaves open the possibility that the administrator might influence the witness's decision (e.g., "Would you like to take another look at number four?"). Even with the best intentions, an administrator might convey subtle cues about which lineup member is the suspect (e.g., by waiting longer for the witness to make a decision about the suspect). Double-blind testing reduces the possibility of the administrator influencing the witness's decision. Some have argued that double-blind testing reduces correct identifications rates (i.e., by minimizing administrator influence that leads to correct identifications; Clark, 2012). However, others have argued (and we agree) that, given witnesses' propensity for identification error, enhancing the reliability of obtained identifications – by ensuring as much as possible that these identification indicate recognition rather than administrator influence – is crucial (Wells, Steblay, & Dysart, 2012).

The use of unbiased instructions (also termed *warning* or *admonishing* the witness) refers to reminding the witness prior to viewing the lineup that the person they are looking for may or may not be present. This simple instruction significantly reduces positive identifications from target-absent lineups (Malpass & Devine, 1981; Steblay, 1997). Omitting this instruction (i.e., using biased instructions) likely increases correct identifications when the perpetrator is present in the lineup (Clark, 2012) but at the great cost of an increase in false identifications from culprit-absent lineups (e.g., Wells, Steblay, & Dysart, 2012).

Simultaneous versus Sequential presentation

Whether to present lineup members all-at-once (simultaneous) or one-at-a-time (sequential) has been a topic of great debate among eyewitness researchers. The idea behind sequential presentation was to reduce the scope for witnesses to compare lineup members in terms of relative similarity to the perpetrator (a *relative* judgment strategy), and encourage witnesses to assess the match between lineup members and their memory of the perpetrator (*absolute* judgment strategy), thus leading to better quality identification decisions (Wells, 1984; Lindsay & Wells, 1985).

Much evidence supports this approach. Meta-analyses of numerous experiments show that, compared to simultaneous presentation, sequential presentation reduces correct identifications from target-present lineups, but reduces incorrect picks from target-absent lineups to a greater extent, resulting in an overall increase in identification accuracy (Steblay, Dysart, Fulero, & Lindsay, 2001; Steblay, Dysart, & Wells, 2011). However, several recent developments have led researchers to question the mechanisms that produce different response patterns for sequential and simultaneous lineups, and the advantages of sequential presentation.

One class of developments concerns new approaches to measuring identification accuracy. Most evidence favouring sequential presentation relies on assessment of response patterns for target-present and -absent lineups, and the diagnosticity of suspect identifications (which speaks to the practical utility of identification responses for informing police investigations of the suspect in question). More recently, some researchers have used signal-detection analyses to assess differences in response patterns, and concluded that these analyses suggest that sequential presentation does not improve witnesses' ability to distinguish perpetrators from innocent lineup fillers, but instead prompts witnesses to be more conservative in their propensity to choose from lineups (Clark, 2012; Palmer & Brewer, 2012). In other words, sequential presentation does not enable witnesses to make better identification decisions; it discourages them from making positive identifications.

Other researchers have used Receiver Operating Characteristic (ROC) analyses to compare responses for sequential and simultaneous lineups. This involves calculating, for each level of identification confidence, the cumulative rate of correct identifications from target-present lineups (correct identification rate) and the cumulative rate of false identifications of an innocent suspect from target-absent lineups (false identification rate) for each level of confidence. Better performance is indicated by a higher ratio of correct identifications to false identifications. Plotting the correct identification rate against the false identification rate for each level of confidence produces a ROC curve; the area under this curve gives an index of overall identification performance.

Comparisons of ROC curves for sequential and simultaneous lineup presentation suggest very different conclusions to those drawn earlier: Sequential presentation is not superior to simultaneous presentation, and in some cases may produce worse identification performance than simultaneous presentation (e.g., Mickes, Flowe, & Wixted, 2012). Some researchers have suggested that simultaneous presentation may produce better identification performance because, compared to sequential presentation, it allows witnesses greater scope to consider diagnostic features when making identification decisions (i.e., features that allow the witness to discriminate between lineup members, as opposed to non-diagnostic features that are shared by all lineup members; Wixted & Mickes, 2014).

One potentially important consideration in comparisons of lineup presentation methods is the role of *backloading* in the sequential lineup. Backloading involves adding extra lineup members in order to conceal from the witness the actual number of people in the lineup (e.g., by adding extra photos to a stack of lineup photos). Without backloading, witnesses shift their decision criterion as they move through a sequential lineup, becoming more likely to make a positive identification as they near the end of the lineup (Horry, Palmer, & Brewer, 2012). Backloading is crucial part of the sequential procedure because it undermines this shift in decision criterion, reducing the likelihood of false identifications from late positions in target-absent sequential lineups (Horry et al., 2012). Although these results show that backloading is important, it has been overlooked in recent comparisons of sequential and simultaneous lineups.

Together, these results paint a somewhat murky picture regarding the superiority of sequential lineup presentation over simultaneous presentation. In our view, however, perhaps the most important point to emerge from the debate about these two procedures is that neither one produces impressive accuracy rates, and neither is likely to prove to be the most effective way of assessing the witness's recognition of the suspect (Brewer & Palmer, 2010; Wells, Memon, & Penrod, 2006).

Novel Approaches to Collecting Identification Evidence

Procedural changes aimed at improving the reliability of identification evidence have generally been conservative. Although research has identified a number of best-practice guidelines for administering lineups that can reduce the risk of a false identification, the nature of the lineup task itself has remained relatively constant: Participants view a series of lineup members and either identify someone as the culprit or reject the lineup as a whole. There have been some variations on this procedure based around the notion of using multiple lineups to better assess witnesses' memory (e.g., Palmer, Brewer, & Weber, 2012; Pryke, Lindsay, Dysart, & Dupuis, 2004; Wells, 1984), but here we consider a departure from this standard practice that suggests a new way of collecting and thinking about identification evidence.

As discussed previously, confidence for recognition memory decisions is thought to index the degree of match between a presented item and an image in memory. With this in mind, researchers have suggested that avoiding explicit categorical identifications and, instead, having witnesses rate their confidence (from 0-100%) that each lineup member is the culprit (referred to as *culprit likelihood ratings*) may provide a number of benefits (Brewer, Weber, Wootton, & Lindsay, 2012; Sauer & Brewer, 2015; Sauer, Brewer, & Weber, 2008a, 2012a; Sauer, Weber, & Brewer, 2012b). First, ratings might provide a more informative index of recognition (i.e., the strength of recognition for the suspect) and discrimination (the extent to which the suspect is favoured over the alternative). Second, compared to categorical responses, ratings may more directly assess the construct of interest: the degree of match between individual lineup members and the witness' memory of the culprit. As suggested above, although an identification probably indicates the selected lineup member is the most plausible of the available options, it says little about how well the selected lineup member matches the witness's memory of the culprit.

Initial tests of this approach provided two encouraging findings. First, research revealed a generally linear, positive relationship between confidence ratings and the likelihood that a face has been previously seen for basic face recognition tasks (Sauer et al., 2012b). Second, after applying algorithms to determine when a rating or pattern of ratings could be taken as indicating a positive identification, ratings were consistently more diagnostic of recognition than categorical responses in basic face recognition and eyewitness identification tasks (Brewer et al., 2012; Sauer et al., 2008a, 2012a; Sauer et al., 2012b).

Collapsing culprit likelihood ratings into categorical classifications allows a demonstration of their diagnostic value, but also reduces the richness of the recognition information provided. Brewer et al. (2012) presented an additional and more informative perspective. For each set of ratings (i.e., for each lineup viewed), the researchers determined whether (a) there was a single highest, maximum rating value and (b) whether the maximum value indicated the suspect. If the maximum value implicated the suspect (i.e., the suspect was favored over the others), the researchers examined variations in the likely guilt of the suspect as a function of the discrepancy between the maximum and next-highest values. This approach produced two notable findings. First, the likely guilt of the suspect increased almost monotonically as a function of the discrepancy between the maximum and next-highest 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%, culprit likelihood ratings were a better predictor of suspect guilt than were categorical identification decisions. Further to demonstrating that patterns of culprit likelihood ratings can offer reliable diagnostic information about suspect guilt for individual witnesses, Brewer et al.'s (2012) findings – specifically the monotonic positive relationship between the discrepancy measure and the likely guilt of the suspect – suggest that a probabilistic treatment of identification evidence may offer a viable alternative to categorical decisions. Relating to the ideological goal of increasing the informational value of identification evidence, the legal system may benefit from eschewing traditional, categorical responses and, instead, considering what patterns of ratings say about the *likely* guilt of the suspect/defendant.

The boundary conditions for ratings-based identification procedures clearly require further investigation. However, such approaches may address a number of systemic problems with traditional identification practices. First, these approaches may attenuate the non-memorial influences on criterion placement that contribute to identification error by compromising 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. Second, these approaches provide legal decision-makers with a richer source of information upon which to base assessments of likely guilt (i.e., speaking to both strength of recognition and degree of discrimination). Further, when a traditional lineup produces a rejection, this provides no information about the degree of match between the suspect and the witness's memory of the culprit (other than that the degree of match did not exceed the criterion for identification). In contrast, in all cases, ratings-based procedures provide 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.

An approach such as this one clearly entails a radical departure from traditional conceptualizations of identification evidence. Lineup tasks would no longer provide a single, categorical outcome that is assumed to somehow resolve the ambiguity around a suspect/defendant's guilt. Instead, triers of fact would need to view a lineup task as providing another source of probabilistic evidence about the possible guilt of the suspect. This may be a difficult notion for police and the courts to accept. However, in response to this concern, two points bear consideration. First, recent research suggests that although mock-jurors might need support in interpreting ratings-based identification evidence, they do not dismiss such evidence as uninformative (Sauer, Palmer, & Brewer, 2017). Second, a radical departure from traditional approaches might be more palatable if one bears in mind how often traditional approaches produce erroneous decisions (e.g., Steblay, et al., 2011).

Novel Approaches to Improving the Accuracy of Child Eyewitness Identifications

Although experiments testing child eyewitness performance tend to be underpowered, the literature provides compelling evidence of a “choosing problem”: Compared to adult participants, children are more likely to falsely identify a suspect from a target-absent lineup (see Fitzgerald & Price, 2015, for a meta-analysis and review). Fitzgerald and Price summarize a number of mechanisms proposed to explain children’s proneness to pick from target-absent lineups. Some explanations revolve around the social demands of the task, and children’s increased susceptibility to suggestion. These explanations suggest that a lineup is, to some extent, inherently suggestive and that children are prone to pick because they believe an identification is expected of them (e.g., Davies, 1996). Other explanations suggest more cognitive mechanisms, proposing that increased false identification rates for child witnesses might reflect under-developed face processing ability (e.g., Davies, Tarrant, & Flin, 1989), an over-reliance on relative familiarity as a cue for decision-making (cf. the ability to recall-to-reject; e.g., Gross & Hayne, 1996), or an inability to effectively process larger stimulus sets (e.g., lineups containing six or more individuals; see Price & Fitzgerald, 2016). A third category of explanations suggests that children’s tendency to pick from target-absent lineups may reflect developmental differences in response inhibition, with younger children experiencing difficulties inhibiting a positive response (e.g., Davies et al., 1989; Zajac & Karageorge, 2009).

Researchers have trialled a variety of approaches, targeting varied combinations of the mechanisms identified above, to address this choosing problem. For example, Pozzulo and Lindsay's (1999) elimination lineup includes a range of procedural elements designed to reduce children's reliance on the relative familiarity of lineup members when making their decision, and help children reject the lineup when the culprit is absent. In the fast-elimination version of this procedure, the child first identifies the lineup member who best matches their memory of the offender and, if the selected lineup member is the suspect, is then explicitly asked if the selected person is the culprit. The fast-elimination version requires the child to begin by eliminating the lineup members who look least like their memory of the culprit. If the final remaining lineup member is the suspect, the child is explicitly asked if that person is the culprit. Either elimination procedure can be combined with modified instructions that emphasize the problem of false identifications and encourage the witness to make an absolute judgment (cf. relying on relative familiarity). Zajac & Karageorge's (2009) "wildcard" technique[1] includes a "tangible rejection option" in the array (e.g., a silhouetted figure or stick drawing) to make the rejection option more salient to child witnesses who may assume they are required to pick from the lineup. This approach also makes the identification and rejection behaviours more similar, because both require the child to actively select an option from the array. Zajac & Karageorge found that, compared to a standard lineup task, the wildcard procedure increased correct rejections (reducing false identifications) from target-absent lineups without reducing correct identifications from target-present lineups (see also Havard & Memon, 2013). Although earlier approaches reported non-significant benefits of similar techniques, these studies showed trends in the expected direction and the failure to reach statistical significance may reflect their reliance on small sample sizes (Ns < 20 per cell; e.g., Beal, Schmitt, & Dekle, 1995; Davies et al., 1989). Finally, Price and Fitzgerald's

[1] Similar approaches have been referred to as the "Mystery" option or "Mr. Nobody" option (see also Beal, Schmitt, & Dekle, 1995; Davies et al., 1989; Havard & Memon, 2013)

(2016) face-off procedure attempts to accommodate children's difficulties with (a) making choices from large arrays and (b) resisting the urge to pick when the target is absent by breaking the lineup task into a series of simple judgments comparing two lineup members.

All three of these approaches represent nice examples of theoretically-motivated procedural innovation, and all three have shown some promise. However, replication across varied stimulus sets and samples, and further investigations of boundary conditions, will be required before any approach can claim decisive empirical support.

Things We Know Little About

Lineup Composition

Researchers (e.g., Wells & Turtle, 1986) and practitioners (e.g., Technical Working Group for Eyewitness Evidence, 1999) generally agree that, when constructing a lineup, a single suspect should be placed among a number of fillers (i.e., known innocent lineup members) selected so that the suspect does not "unduly stand out" as the only plausible candidate for identification. Essentially, the lineup should be "fair". However, intuitive as this proposal is, it is unclear how best to achieve this goal. Here we consider some of the unresolved questions relating to lineup composition and lineup fairness. We identify three broad areas in which the literature falls short of providing clear and compelling guidance. First, how many fillers should be in a lineup? Second, how should these fillers be selected? Third, how generalizable are lab findings to applied settings?

Lineup size. Given that a lineup should contain one suspect and some fillers, an obvious question is how many fillers are required? Across jurisdictions, there is considerable variation in requirements (or guidelines) relating to *nominal size* (i.e., the number of people in the lineup). For example, the United Kingdom's VIPER system presents nine member lineups (i.e., one suspect and eight fillers) while, in the US, the National Institute of Justice's (1999) guidelines on eyewitness identification protocol suggest "a minimum of five fillers" for a photo-array and four for a live lineup. Canada's Sophonow Inquiry Report recommended including at least 9 fillers, while Russia's Criminal-Procedure Code suggests a minimum of 2 fillers. Given this variability, it may be of some comfort that, although empirical investigations of the effects of nominal size are scarce, there is some evidence that nominal size has little effect on correct or false identification rates (Nosworthy & Lindsay, 1990; although, very large arrays are likley to reduce false identifications of innocent suspects based on guesses; Levi, 1998). A more important consideration is likely to be the *functional size* of the lineup (the number of plausible candidates in the lineup). However, despite a general consensus that functional size is important and that low functional size increases the risk of false identification, this issue remains under-researched. Although the literature suggests that a functional size of three or more represents a fair lineup (Brigham, Ready, & Spier, 1990; Nosworthy & Lindsay, 1990), the boundary conditions for this suggestion remain largely untested. For example, we do not know how the effects of functional size are moderated by factors affecting memory quality, or lineup presentation, or influences on witness's decision criteria.

Selecting fillers. We now consider what we know, and do not know, about selecting fillers for a lineup. The over-arching and intuitive principle is that selection of fillers should promote lineup fairness. However, fairness is multi-faceted. First, the lineup should be fair for the suspect; ensuring that the suspect does not unduly stand out. Thus, there must be some degree of physical similarity between the suspect and the fillers. Second, the lineup should be fair for the witness. Luus and Wells (1991) argued that a lineup in which the degree of similarity between the suspect and fillers is too high places an unreasonable demand on a witness's memory and capacity for discrimination. Essentially, then, a fair lineup must offer some protection to innocent suspects while still allowing a witness to recognize and identify a guilty suspect. Two approaches to selecting fillers are commonly discussed in the literature. The match-to-suspect (or similarity) approach involves selecting fillers based on their physical similarity to the suspect. As an alternative, Luus and Wells (1991) proposed the match-to-description approach, arguing that selected fillers should possess all the physical characteristics included in the witness's description of the culprit but may vary on any non-described features. Theoretically, the match-to-description approach has a number of advantages over the match-to-similarity approach. First, it ensures that fillers possess all the features salient enough to be included in the witness's recalled memory for the perpetrator. In contrast, a match-to-suspect approach may include fillers who do not possess features that were salient to the witness, if these features were not also salient to the person constructing the lineup. Second, it allows for some heterogeneity among lineup members to aid the witness's discrimination, whereas a match-to-similarity may place an unreasonable demand on the witness's ability to discriminate. Finally, compared to subjective perceptions of physical similarity, the description provides a more objective and concrete basis for filler selection. These sound theoretical grounds for recommending the match-to-description approach may explain why this approach has received fairly consistent "in-principle" support in the literature (e.g., Wells et al., 1998), and been included in the National Institute of Justice guidelines (1999).

Interestingly, however, empirical support for the match-to-description approach is limited. Although Wells, Rydell, and Seelau (1993) found that the match-to-description approach improved correct identifications rates (with no increase in false identification rates), subsequent research revealed either non-significant differences in identification accuracy rates or, in one case, that the match-to-description approach increased false identification rates along with gains in correct identifications (see Fitzgerald, Oriet, & Price, 2015, for a review). Although power issues in individual studies and differences in description quality across studies may account for some of the variability in findings, meta-analyses suggest that, overall, filler selection strategy has little effect on the diagnosticity of suspect identifications (Clark & Godfrey, 2009; Clark, Howell, & Davey, 2008).

Moreover, there are potentially important practical limitations to the match-to-description approach. First, it relies heavily on the quality of the witness's initial description and the limited available research suggests that, especially in the field, witnesses' descriptions omit critical details. For example, according to Lindsay, Martin, and Webber's (1994) data, less than half of real-crime witnesses' descriptions included details on age, race, height, build, hair color or length. Less than 10% of descriptions included details on facial features (e.g., eyes, complexion, or facial hair). Second, witnesses will probably not report all recalled details: Some details, even if distinctive and/or vividly recalled, may be difficult to articulate. Thus, the descriptions used as a basis for filler selection may omit important information required, or fail to provide sufficient information, to construct a fair lineup, and it seems likely that some combination of the match-to-description and match-to-suspect approaches will be required. Indeed, the National Institute of Justice guidelines (1999) suggest that when a description provides an inadequate basis for selecting fillers, selection should be based on ensuring fillers "resemble the suspect in significant features" (p.29). However, it is not clear how the significance of any such features should be determined.

Despite these issues with match-to-description lineups, ensuring that, at a minimum, all lineup members match the witness's description of the culprit provides an easy and effective method for preventing the presentation of a severely biased lineup in which only the suspect stands out. Once a pool of fillers is identified that meets that criterion, filler selections can be prioritized according to their similarity to the suspect. Unfortunately, however, we do not yet have any objective guidelines or procedures for similarity-based filler selection that will maximize the probability of a correct identification and minimize the likelihood of a mistaken identification.

Generalizing From the Lab to Applied Settings

Applied research always entails a compromise between experimental control and ecological validity. Design choices that justifiably promote experimental control will often necessitate important caveats on the generalizability of findings to applied settings. Here we consider an important example of this issue relevant to understanding lineup composition effects on identification performance. When constructing lineups in lab environments, researchers typically select lineup fillers from a pool of potential lineup members based on some combination of the potential fillers' match to (a) a description of the target (often obtained from a group of pilot participants) and (b) the target's physical appearance (typically assessed through a blend of rigorous visual inspection and intuition). The target-present lineup is then constructed by placing the target amongst the selected fillers. Typically, the target-absent lineup keeps the same fillers, and replaces the target with a designated innocent suspect selected from the original pool of fillers, and generally bearing a relatively high resemblance to the target. This "same fillers" approach allows for greater experimental control when examining the effects of experimental manipulations on correct and false identification rates. However, it differs importantly from the filler selection process in applied settings (see Clark & Tunnicliff, 2001, for a review). In applied settings, investigators often do not know who the actual target is. Thus, in target-absent lineups, fillers are selected based on their match to the *suspect*, not their match to the *target*. Clark and Tunnicliff compared false identification rates using the match-to-target and match-to-suspect approaches to filler selection for target-absent lineups. Compared to the match-to-suspect approach (common in applied settings), the match-to-target approach (common in lab settings) produced lower false identification rates, and a lower conditional probability of innocent suspect identification (i.e., the likelihood the innocent suspect was identified, given the witness picked someone from a target-absent lineup). Thus, the typical experimental approach to selecting target-absent fillers may underestimate innocent suspect identification rates in applied settings. Consequently, to the limited extent that the literature does speak to lineup composition effects on identification performance in lab settings, we should exercise caution when generalizing these findings to real-world identification performance.

In sum, although few would argue that lineup composition is an important factor in understanding identification performance – and that biased lineups increase the risk of false identifications – a careful consideration of the literature reveals that we still know relatively little about how best to construct fair lineups for applied settings. Brewer, Weber, and Semmler (2005) noted that the absence of a thorough and systematic body of literature investigating lineup composition effects on responding has prevented a clear understanding of many important underlying issues. Unfortunately, this remains the case.

Influence of Non-memorial Cues

One potentially important class of variables are non-memorial factors that influence identification decisions. Ideally, witnesses would make identification decisions based only on the degree of match between their memory of the perpetrator and the members of the lineup (e.g., Brewer et al., 2012; Sauer et al., 2008a). However, this is not always the case; witnesses often pick someone from a lineup even when the match between their memory and the chosen lineup member does not support such a decision (e.g., Wells, 1993). This can happen, for example, if the witness assumes that the actual perpetrator is in the lineup and that their task is to pick out that person.

In situations where a witness is motivated to choose someone from a lineup but cannot do so on the basis of a match with their memory of the perpetrator, factors unrelated to memory can influence identification decisions. This notion aligns well with various decision making models (e.g., Gigerenzer & Goldstein, 1996; Kahneman & Frederick, 2002). According to such models, when faced with a decision people attempt to make a good decision based on directly relevant information. However, if a good decision is unable to be made based on such information, the decision maker might turn to other cues that are perceived as valid indicators that might support a decision. In the present context, an eyewitness should attempt to base their identification decision on the degree of match between their memory of the perpetrator and members of the lineup. However, if a decision cannot be reached this way, then the witness might consider other, non-memorial cues. Perhaps the most obvious possibility is that the witness could look for cues from the lineup administrator that could indicate which lineup member is the suspect. However, even in the absence of such cues (e.g., if the lineup is administered via computer), various non-memorial cues can influence identification decisions.

For example, the witness might look at cues relating to facial expression or body language. If, for example, one lineup member is perceived as less trustworthy, more stereotypically “criminal” in appearance, or more nervous than others, it could provide the basis for choosing that person from the lineup (e.g., Flowe, 2012; Flowe, Klatt, & Colloff, 2014; Weigold & Wentura, 2004; Wilson & Rule, 2015). From an objective viewpoint, such cues are clearly not reliable, diagnostic indicators of the identity of the perpetrator. However, if the witness does not dismiss such cues as non-valid, they can provide a basis for choosing one lineup member over the others.

Identification decisions might also be influenced by non-memorial cues unrelated to the perceived nature of lineup members, such as variations in the quality of photographs (e.g., does one image appear clearer than others or with a different background?). Alternatively, a witness might consider where in the lineup the suspect is most likely to appear. Systematic positon biases occur in many hide-and-seek tasks (e.g., games of “battleships”; students guessing on multiple-choice questions; Bar-Hillel, 2015), and similar biases influence guessing from lineups. Witnesses expect suspects to be more likely to appear in central locations in simultaneous lineups rather than edge locations, and the top row (rather than bottom row) of a photo-array arranged as a grid (O’Connell & Synnott, 2009; Palmer, Sauer, & Holt, 2017). Position effects also occur in sequential lineups. Choosing increases as the lineup progresses and, as a result, the target (or a particularly plausible filler) is increasingly likely to be identified from later positions in the lineup (e.g., Carlson, Gronlund, & Clark, 2008; Clark & Davey, 2005; Horry, Palmer, & Brewer, 2012).

Overall, the effect of non-memorial cues on identification decisions should be modest, because we assume that most witnesses will try to base their decision on an assessment of memory match. However, there may be specific circumstances under which such cues have a much larger effect. For example, if a witness feels a strong sense of expectation to pick someone from the lineup (e.g., “we’re really keen to nail this guy”) and feels that all lineup members are plausible matches for the suspect, the influence of non-memorial cues that allow one lineup member to be distinguished from the others might be especially strong.

Influence of Memorial Factors

There are likely to be some factors that are not well understood but likely play an important role in eyewitness identification decisions. Our limited understanding of such factors may reflect a paucity of data on the issue, the effect being difficult to capture in laboratory studies, or the fact that the effect emerges only under certain conditions.

Context reinstatement. One example is context reinstatement; that is, re-establishing the context from encoding at the time of a memory test. Context reinstatement has proven to be an effective means of increasing the recall of accurate information from memory (e.g., Godden & Baddeley) and for this reason is included in the cognitive interview protocol (Fisher & Geiselman, 1992). However, its influence on recognition memory tasks is less clear-cut (Smith, Glenberg, & Bjork, 1978; Smith & Vela, 2001). Similarly, effects of context reinstatement on eyewitness identification accuracy have produced varying results. Some studies have shown benefits of reinstatement for identification accuracy and others have not, although the data overall suggest that identification accuracy may benefit from physical context reinstatement (i.e., actually returning to the encoding context) more than mental reinstatement (e.g., Cutler & Penrod, 1988; Cutler, Penrod, & Martens, 1987b; Gwyer & Clifford, 1997; Krafka & Penrod, 1985; Sanders, 1984; Smith & Vela, 1992).

Unconscious transference. Another example is unconscious transference, whereby a witness confuses an innocent person (e.g., a bystander at the scene of a crime) for the perpetrator they are trying to recognize (e.g., Loftus, 1976). This can occur if a witness incorrectly attributes the familiarity of the innocent person with the context of having seen that person commit the crime in question (e.g., Perfect & Harris, 2003) or if the witness fails to notice that the perpetrator and the innocent bystander are not the same person (e.g., Davis, Loftus, Vanous, & Cucciare, 2008; Fitzgerald, Oriet, & Price, 2014). Although unconscious transference effects have been found in some studies, they have not emerged in others. Perhaps most notably, one series of highly realistic field experiments conducted under highly realistic conditions (i.e., numerous perpetrators and bystanders; retention intervals varying up to 2 weeks) failed to produce any evidence of unconscious transference (Read, Tollestrup, Hammersley, McFazden, & Christensen, 1990). This highlights the importance of considering conditions that might facilitate the mechanisms thought to underpin unconscious transference. Even if clear and reliable effects do not emerge across experiments, very strong effects may occur in isolated cases when the requisite conditions align.

The Importance of Considering Interactions between Variables

Throughout this chapter, we have touched on the importance of examining interactions between factors that influence eyewitness identification decisions. We believe this is a crucial issue for future research in this field, because there are many theoretically-motivated reasons why the effects of one factor might be expected to vary depending on some other factor. For example, retention interval influences identification performance, but its effects may be moderated by numerous other factors that promote the formation of a strong memory at encoding. These might include the distinctiveness of the perpetrator: that is, shorter retention intervals promote better memory performance, but the benefits of a short delay between encoding and identification tests may be smaller when the perpetrator is especially distinctive in appearance (and, hence, memorable even after a long delay).

Exposure duration might prove a consistent moderator of the effects of other factors known to affect identification accuracy. For example, the presence of a weapon can impair memory for a perpetrator because it draws attention away from the appearance of the perpetrator. However, this effect might diminish with longer exposure duration. That is, if the witness views an armed perpetrator for an extended time, the witness might begin to direct more attention to the perpetrator’s appearance, especially if the witness’s perception of physical threat diminishes over time. Similarly, cross-race effects (whereby identification accuracy is worse when attempting to identify someone from another race rather than one’s own race) might weaken under long exposure duration conditions if exposure is sufficient to facilitate a strong memory for an other-race perpetrator.

The need to study potential interactions between variables that affect eyewitness identification clearly presents a challenge to eyewitness researchers. It also highlights the fact that, although research in this area has progressed enormously over the past few decades, there is still much work to be done.

The Base Rate for Culprit Present Lineups

Finally, we highlight one generally neglected example of how the full applied implications of much eyewitness identification research can only be appreciated by recognizing the match, or mismatch, between one aspect of the researcher’s experimental design and the reality of lineup administration practices in the real world. A common practice in laboratory identification research is to present witnesses with either culprit-present or absent lineups, with studies often presenting 50% of each. One of the reasons for following this approach is obviously to allow researchers to distinguish whether experimentally manipulated variables are affecting witnesses’ tendencies to choose from the lineup or the capacity to make accurate recognition decisions. As Brewer et al. (2005) noted, however, conclusions about precisely how individual variables affect identification performance is likely to vary depending on the base rate of culprit-present lineups. In a similar vein, Brewer and Wells (2006, p.25) showed how varying the proportion of culprit-present lineups leads to quite different conclusions about the relationship between identification confidence and accuracy and, specifically, about the diagnostic value of confidence for determining accuracy. More recently, Wells, Yang, and Smalarz (2015) conducted the first comprehensive examination of how culprit-present base rates can influence the reliability of identification test evidence under a variety of different experimental manipulations. Of course, as noted by both Brewer et al. (2005) and Wells et al. (2015), culprit-present base rates likely vary markedly depending on idiosyncratic jurisdictional practices (e.g., base rates will probably be lower if police use the lineup as a hypothesis testing instrument rather than as the culmination of an exhaustive investigative process), with such variations having important implications for the information gained from positive identifications, lineup rejections, and filler picks (cf. Wells et al., 2015).

Conclusion

Although researchers have developed many improvements to identification procedures, it is vital to remember that implementing these in police settings may not be straightforward. Consider the example of double-blind lineup administration, a procedure supported by strong evidence and endorsed widely by researchers. This procedure requires that the officer investigating the case (with knowledge of which lineup member is the suspect) does not conduct the linuep. Although the benefit of this policy might seem obvious to a researcher, an officer who has invested time and effort into building rapport with a witness may be reluctant to hand over a crucial component of the investigation to another person. The uptake of recommended procedures among investigators can be facilitated via carefully developed training protocols that (a) explain the evidence behind procedures and the benefits of implementing procedures, and (b) implement them in minimally disruptive ways. For example, double-blind administration allows prosecutors at trial to rule out undue influence on the witness from the lineup administrator. This could be done with the investigating officer retaining control of all aspects of the investigation and only being absent for the actual presentation of the lineup, which would be administered by another officer or via computer (e.g., Brooks III, 2017; Norwood Police Department, 2017). In a similar vein, obtaining an immediate post-decision record of the witness's confidence would be very simple to achieve using computerized testing. The final chapter of this volume provides a detailed discussion of challenges and issues relevant to translating research findings into policy.

References

Bar-Hillel, M. (2015). Position effects in choice from simultaneous displays: A conundrum solved. *Perspectives on Psychological Science*, *10*, 419-433.

Beal, C. R., Schmitt, K. L., & Dekle, D. J. (1995). Eyewitness identification of children: Effects of absolute judgments, nonverbal response options, and event encoding *Law & Human Behavior, 19*, 197-216.

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.

Brewer, N. (2011). The Police Journal (Opinion section): *Practical advantages in computerized photo line-ups*. (April 2011 Volume 92) http://journal.pasa.asn.au/

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and 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. & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15*, 77-96.

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology, 22*, 827-840.

Brewer, N., Weber, N., & Semmler, C. (2005). Eyewitness Identification. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective.* (pp. 177-221). New York: Guilford.

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science*, *23,* 1208-1214.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, functional size and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30.

Brewer, N., & Wells, G. L. (2009). Obtaining and interpreting eyewitness identification test evidence: The influence of police-witness interactions. In T. Williamson, R. Bull, & T. Valentine (Eds.), *Handbook of psychology of invetigative interviewing: Current developments and future directions* (pp. 205-220). Chichester: Wiley-Blackwell.

Brigham, J. C., Ready, D. J., & Spier, S. A. (1990). Standards for evaluating the fairness of photograph lineups. *Basic and Applied Social Psychology, 11*, 149-163.

Brooks III, W. G. (2017, May 24). *US identification procedures*. Eyewitness Identification Workshop. Boston, May 2017.

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, 14, 118-128.

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, *7*, 238–259.

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29*, 151-172.

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review, 16*, 22-42.

Clark, S. E., Howell, R., & Davey, S. (2008). Regularities in Eyewitness Identification. *Law and Human Behavior, 32*, 187-218.

Clark, S. E., & Tunnicliff, J. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behavior, 25*, 199-216.

Cutler, B. L., & Penrod, S. D. (1988). Context reinstatement and eyewitness identification. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 231-244). Oxford, England: John Wiley.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior*, *14*, 185-191.

Davis, D., Loftus, E., Vanous, S., & Cucciare, M. (2008). Unconscious transference can be an instance of change blindness. *Applied Cognitive Psychology*, *22*, 605-623.

Davies, G. (1996). Children's identification evidence. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 233–258). Mahwah, NJ: Erlbaum.

Davies, G., Tarrant, A., & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology, 80,* 415.

Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior, 6,* 15-30.

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology, 87*, 951-962.

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67*, 818-835.

Fisher, R. P. & Geiselman, R. E. (1992). Memory-enhancing techniques in investigative interviewing: The cognitive interview. Springfield, IL: C.C. Thomas.

Fitzgerald, R. J., Oriet, C. & Price, H. L. (2014). Change blindness and eyewitness identification: Effects on accuracy and confidence. *Legal and Criminological Psychology*, *21*, 189-201.

Fitzgerald, R. J., Oriet, C., & Price, H. L. (2015). Suspect filler similarity in eyewitness lineups: A literature review and a novel methodology. *Law & Human Behavior, 39*, 62-74.

Fitzgerald, R. J., & Price, H. L. (2015). Eyewitness identification across the life span: A meta-analysis of age differences. *Psychological Bulletin*, *141*, 1228-1265.

Flowe, H. D. (2012). Do characteristics of faces that convey trustworthiness and dominance underlie perceptions of criminality? *PLoS ONE*, 7: e37253.

Flowe, H. D., Klatt, T., & Colloff, M. F. (2014). Selecting fillers on emotional appearance improves lineup identification accuracy. Law and Human Behavior, 38, 509-519.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650-669.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*, 325–331.

Gross, J., & Hayne, H. (1996). Eyewitness Identification by 5- to 6-Year-Old Children. *Law & Human Behavior*, *20*, 359-373.

Gwyer, P. and 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.

Havard, C., & Memon, A. (2013). The Mystery Man Can Help Reduce False Identification for Child Witnesses: Evidence from Video Line-ups. *Applied Cognitive Psychology*, *27*, 50-59.

Horry, R., & Brewer, N. (2016). How target-lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology: General*, *145*, 1615-1634.

Horry, R., Palmer, M. & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied, 18*, 346-360.

Innocence Project. (2017). *Innocence Project*. Retrieved December 1, 2017, from http://www.innocenceproject.org/about/index.php

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.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp.49-81). New York: Cambridge University Press.

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.

Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology, 97*, 286-314.

Kovera, M. B., & Evelo, A. J. (2017). The case for double-blind lineup administration. *Psychology, Public Policy, and Law, 23*, 421-437.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology, 49*, 58-69.

Levi, A. M. (1998). Protecting innocent defendants, nailing the guilty: A modified sequential lineup. *Applied Cognitive Psychology, 12*, 265-275.

Light, L. L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 212-228.

Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default Values in Eyewitness Descriptions: A Problem for the Match-to-Description Lineup Foil Selection Strategy. *Law & Human Behavior, 18*, 527-541.

Lindsay, R. C. L., Semmler, C., Weber, N., Brewer, N., & Lindsay, M. R. (2008). How variations in distance affect eyewitness reports and identification accuracy. *Law and Human Behavior, 32*, 526-535.

Loftus, E. (1976). Unconscious transference in eyewitness identification. *Law and Psychology Review*, *2*, 93-98.

Luus, E. C. A., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law & Human Behavior*, *15*, 43-57.

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482-489.

Mansour,J. K. & Flowe, H. D. (2010). Eyetracking and eyewitness memory. *Forensic Update* No. 101, Autumn, 2010.

Mansour, J. K., Lindsay, R. C. L., Brewer, N., & Munhall, K. G. (2009). Characterizing visual behaviour in a lineup task. *Applied Cognitive Psychology*, *23*, 1012-1026.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces - A meta-analytic review. *Psychology, Public Policy, and Law*, *7*, 3-35.

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, *94*, 339-354.

Norwood Police Department. (2017). Eyewitness identification 2017 (Policy and Procedure No. 1.12). Retrieved 24 May 2017 from: http://www.norwoodma.gov/departments/police/mptc_training_material.php

Nosworthy, G. J., & Lindsay, R. C. L. (1990). Does Nominal Lineup Size Matter? *Journal of Applied Psychology*, *75*, 358-361.

O'Connell, M., & Synnott, J. (2009). A position of influence: Variation in offender identification rates by location in a lineup. *Journal of Investigative Psychology and Offender Profiling*, *6*, 139-149.

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica, 133*, 137-145.

Palmer, M. A., Brewer, N., & Weber, N. (2012). The information gained from witnesses' responses to an initial "blank" lineup. *Law and Human Behavior*, *36*, 439–4 47.

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied, 19*, 55-71.

Palmer, M. A., Sauer, J. D., & Holt, G. A. (2017). Undermining position effects in choices from arrays, with implications for police lineups. *Journal of Experimental Psychology: Applied*, *23*, 71-84.

Perfect, T. J., & Harris, L. J. (2003). Adult age differences in unconscious transference: Source confusion or identity blending? *Memory and Cogni*tion, *31*, 570-580.

Potter, R., & Brewer, N. (1999). Perceptions of witness behaviour accuracy relationships held by police, lawyers and jurors. *Psychiatry, Psychology and Law*, *6*, 97-103.

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, *84*, 167-176.

Price, H. L., & Fitzgerald, R. J. (2016). Face-Off: A New Identification Procedure for Child Eyewitnesses. *Journal of Experimental Psychology: Applied, 22*, 366-380.

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: A method of calibrating eyewitness identifications. *Journal of Applied Psychology*, *89*, 73-84.

Read, J. D., Tollestrup, P., Hammersley, R., McFadzen, E. and Christensen, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology*, *4*, 3–31.

Sanders, G. S. (1984). Effetcs of context cues on eyewitness identificaiton responses. *Journal of Applied Social Psychology*, *14*, 386-397.

Sauer, J. D., & Brewer, N. (2015). Confidence and accuracy of eyewitness identification. In T. Valentine & J. P. Davis (Eds.), *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV* (pp. 185-208). Chichester: Wiley Blackwell.

Sauer, J. D., Brewer, N., & Weber, N. (2008a). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, *137*, 528-547.

Sauer, J. D., Brewer, N., & Weber, N. (2012a). Using confidence ratings to identify a target among foils. *Journal of Applied Research in Memory and Cognition, 1*, 80-88.

Sauer, J. D., Brewer, N., & Wells, G. L. (2008). Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Legal and Criminological Psychology*, *13*, 123-135.

Sauer, J. D., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Human Behavior, 34*, 337-347.

Sauer, J. D., Palmer, M. A., & Brewer, N. (2017). Mock-juror evaluations of traditional and ratings-based eyewitness identification evidence. *Law & Human Behavior, 41*, 375-384.

Sauer, J. D., Weber, N., & Brewer, N. (2012b). Using ecphoric confidence ratings to discriminate seen from unseen faces: The effects of retention interval and distinctiveness. *Psychonomic Bulletin & Review, 19*, 490-498.

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied, 15*, 46-62.

Semmler, C., & Brewer, N. (2006). Postidentification feedback effects on face recognition confidence: Evidence for metacognitive influences. *Applied Cognitive Psychology*, *20*, 895-916.

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.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology, 85*, 542-550.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353.

Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology*, *6*, 125-139.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*, 203–220.

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. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D.F. Ross, D.J. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300-327). New York: Cambridge University Press.

Sporer, S. L., Penrod, S. D., Read, D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*, 315-327.

Steblay, N. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413-424.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*, 283-297.

Steblay, N. M., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, *25*, 459-473.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showups and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27*, 523-540.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17*, 99-139.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington DC: U.S. Department of Justice, Office of Justice Programs.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal of Experimental Psychology: Applied*, *10*, 139-147.

Weigold, A., & Wentura, D. (2004). Who's the one in trouble? Experimental evidence for a 'psychic state' bias in lineups. *European Journal of Social Psychology*, *34*, 121-133.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology*, *36*, 1546-1557.

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, *14*, 89-103.

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571.

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., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science*, *10*, 138-144.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest, 7,* 45-75.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 274-295.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78,* 835-844.

Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. Journal of Applied Research in Memory and Cognition, 4, 313-317.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photo spreads. *Law and Human Behavior, 22*, 603-647.

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2012). Eyewitness identification reforms: Are suggestiveness-induced hits and guesses true hits? *Perspectives on Psychological Science, 7*, 264-271.

Wells, G. L., & Turtle, J. W. (1986). Eyewitness Identification: The Importance of Lineup Models. *Psychological Bulletin, 99*, 320-329.

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior, 39*, 99-122.

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychological Science, 26*, 1325-1331.

Wixted, J. T. & Mickes, L. (2014). A signal-detection-based diagnostic feature-detection model of eyewitness identification. *Psychological Review, 121*, 262-276.

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences, 113,* 304-309.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*, 10-65.

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology, 23*, 358-368.