Eyewitness identification tests

Neil Brewer* and Matthew A. Palmer
School of Psychology, Flinders University, Adelaide, Australia

When investigating crimes police frequently use eyewitness identification tests such as lineups in an attempt to establish the culprit's identity. It is now well documented that witnesses commonly make mistakes, sometimes identifying innocent suspects or failing to identify the culprit. Lineup administrators also make mistakes in the sense that they use procedures that increase the likelihood of eyewitness error. Eyewitness memory research has provided a basis for many useful guidelines regarding the conduct of identification tests. Here, we review the relevant research underpinning effective procedures for lineup construction, lineup presentation, recording the witness's decision, and providing feedback to the witness. The systematic implementation of these guidelines is likely to reduce significantly the extent of eyewitness errors at the identification test and improve the probative value of courtroom testimony about identity. However, further significant advances may well require the development of novel procedures for accessing witnesses' memories.

Identification tests - or lineups - are an integral part of police investigative procedures. Police will often ask eyewitnesses to crimes to view a lineup or photoarray in an attempt to test the validity of their hypotheses about an offender's identity or to strengthen their evidence against a suspect. Sometimes the witness's decision about the lineup may be the key evidence against a suspect. Even when it only constitutes part of the evidence, it is well established that identification evidence is particularly persuasive for those assessing the likely guilt of a defendant (Cutler & Penrod, 1995; Wright, 2007). Yet, there now exists abundant evidence illustrating the fallibility of witnesses. Witnesses sometimes mistakenly identify innocent suspects, often with tragic consequences for the person identified (Cutler & Penrod, 1995; Innocence Project, 2008). They also sometimes identify a lineup foil (i.e. a known-innocent lineup member), perhaps delaying or even thwarting a productive police investigation or perhaps resulting in the rest of their testimony being given little weight. And, witnesses often fail to pick anyone out from the lineup (Pike, Brace, & Kynan, 2002), making it difficult for the police to secure a successful prosecution if the suspect is indeed the culprit. These 'errors' can occur even when there are multiple witnesses. So, whenever the identity of a person is crucial in a police investigation and subsequent trial, the result of an identification test is likely to be an important component of the evidence.

*Correspondence should be addressed to Dr Neil Brewer, School of Psychology, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia (e-mail: neil.brewer@flinders.edu.au).

DOI:10.1348/135532509X414765
Just as witnesses can make mistakes when confronted with a lineup, so too can the police or other officials administering the lineup. Here, we will examine some of the procedures associated with running a lineup that increase the likelihood of errors being made by witnesses and diminish the quality of courtroom testimony about identity. We outline the current state of knowledge on how to minimize the possibility of error, and improve the quality of identification evidence. One of the important general observations that we make from the outset is that the identification test, no matter how well conducted, can only provide so much. Because person identification is so crucial in many investigations and trials, it is perhaps tempting to believe that identification tests can be structured and managed so as to guarantee that a witness’s decision is definitive regarding the guilt or innocence of the suspect. However, while eyewitness memory research provides a number of valuable pointers for conducting identification tests, this belief is simply not justified. Our present state of knowledge is such that, even if all guidelines suggested by empirical research were followed appropriately and consistently, eyewitnesses will still sometimes pick innocent suspects and known-innocent lineup foils, and will reject lineups containing the culprit. On the bright side, given that research to date has only scratched the surface in terms of the permutations of variables examined, we are hopeful that future work will gradually help us refine these guidelines.

Here, we review what we know about a number of the effective steps in conducting a lineup. We cover (1) lineup construction; (2) lineup presentation; (3) recording the witness’s decision; and (4) feedback to the witness. Note that in various jurisdictions guidelines for lineup conduct already exist, often shaped by empirical research (e.g. Technical Working Group for Eyewitness Evidence, 1999; UK Home Office, 2008). While our review’s conclusions may sometimes correspond with aspects of such guidelines, we use as our starting-point what we consider to be reliable bodies of empirical research rather than current policies and procedures.

First, however, we provide a brief overview of some basic identification test jargon. Lineups can either be culprit (target)-present (i.e. the actual culprit is in the lineup) or -absent (i.e. an innocent suspect, not the actual culprit, is in the lineup). The witness can either make a positive or negative identification decision: that is, they can make a lineup choice (chooser) or they can reject the lineup (non-chooser). Positive decisions include (i) a correct identification of the culprit from a culprit-present lineup, (ii) a false identification of an innocent suspect from a culprit-absent lineup, and (iii) a mistaken or incorrect identification of a (known-innocent) lineup foil or filler. Negative decisions include (i) a correct lineup rejection where the witness correctly rejects (i.e. fails to choose from) a culprit-absent lineup, and (ii) an incorrect lineup rejection where the witness rejects (i.e. fails to choose from) a culprit-present lineup.

**Lineup construction**

Here, we discuss some important issues regarding the construction of lineups, including the number of suspects in the lineup, criteria for the selection of lineup foils, and the actual and functional size of lineups.

**Number of suspects**

The number of suspects in a lineup can range from one (i.e. a single-suspect lineup) – where one lineup member is the suspect and all other lineup members are
known-innocent foils or fillers - to every member of the lineup (i.e. an all-suspect lineup). Both types of lineup are used in real investigations, along with hybrid models containing some suspects and some foils (e.g. Wells & Turtle, 1986). Wells and Turtle have argued convincingly that single-suspect lineups yield lower false identification rates than do all-suspect lineups. Single-suspect lineups allow positive identifications to be classified as suspect identifications or known errors, with only the former potentially leading to prosecution. In contrast, any positive identification from an all-suspect lineup will likely result in prosecution. This is problematic, given the overwhelming evidence that witnesses often choose innocent people from lineups (e.g. Steblay, Dysart, Fulero, & Lindsay, 2001; Wells, 1993). Indeed, in several archival studies of real cases, approximately one third of all positive identifications were of known foils (see Wells, Memon, & Penrod, 2006). If all-suspect lineups were standard practice, this would represent a substantial number of false identifications and, hence, a substantial number of wrongful prosecutions.

Proponents of all-suspect lineups might argue (probably correctly) that increasing the number of suspects should also increase the chances that the true culprit will appear in the lineup, thereby reducing the likelihood of error. However, Wells and Turtle’s (1986) analysis of one data set demonstrated that even if an all-suspect lineup could ensure the presence of the culprit, a single-suspect lineup would only need to contain the actual culprit more than one third of the time to yield a lower false identification rate than the all-suspect lineup. Thus, any gains achieved by increasing the likelihood of the culprit being in the lineup would almost certainly be outweighed by a sizable increase in the chance of a false identification of an innocent suspect.

Selection of lineup foils
Another important factor to consider in constructing lineups is the criterion for selection of foils. The characteristics of foils can markedly influence the extent to which a suspect stands out in the lineup and, hence, the pattern of identification responses. In extreme cases, lineups may contain only one plausible match for the culprit. For example, a black suspect may be placed in a lineup of white foils, as has happened in at least one real case (Ellison & Buckhout, 1981). A less striking, but no less problematic, example might involve placing a suspect with bright blue eyes amongst a group of similar looking foils, but all with dark brown eyes, when the witness has previously described the culprit as having distinctive bright blue eyes.

These examples illustrate the distinction between the nominal size and functional size of a lineup (Wells, Leippe, & Ostrom, 1979). Nominal size simply refers to the number of lineup members, whereas functional size refers to the number of lineup members who provide a viable option for selection. In the first of the two previous examples, regardless of how many white foils were included (i.e. regardless of the nominal size of the lineup), the functional size would be one. Likewise, if the witness clearly recalled the culprit having bright blue eyes, the functional size of the second lineup would also be one.

There is evidence suggesting that, provided there are at least three plausible foils, the nominal size of lineups does not influence choosing, correct identification, or false identification rates (Nosworthy & Lindsay, 1990). However, although the nominal size of lineups appears to have little effect, the functional size of lineups is important. Not surprisingly, low functional size has been shown to increase false identification rates
from culprit-absent lineups (e.g. Lindsay & Wells, 1980; Wells, Rydell, & Seelau, 1993; but cf. Gonzalez, Ellsworth, & Pembroke, 1993). In principle, enhancing the fairness of lineups may seem relatively uncomplicated; provided a sufficient number of plausible foils are included, a lineup should be fair. However, developing a method for the selection of plausible foils has not been straightforward.

A common strategy is to choose foils on the basis of physical similarity to the suspect. However, this approach often relies on subjective judgment of the person constructing the lineup and, hence, there is the potential for substantial variability in what constitutes a desirable degree of similarity. A more specific problem is that this strategy can lead to the selection of foils bearing too much resemblance to the suspect which, in turn, will likely reduce accuracy (Tredoux, 2002; Wells et al., 1993). To illustrate this point, Luus and Wells (1991) pointed out that a hypothetical lineup of clones would make it impossible for a witness to make a meaningful decision. Thus, the problem is one of how to provide an appropriate test of a witnesses’ memory for the culprit, without impeding the ability of a good witness to identify the culprit (Luus & Wells, 1991). Interestingly, although a reduction in correct identifications is the most obvious grounds for criticism of similarity-based foil selection, there is some suggestion that this approach can also increase false identifications from culprit-absent lineups (e.g. Wolgater, Marwitz, & Leonard, 1992).

Another strategy is to select foils that match the description of the culprit (Luus & Wells, 1991). There is evidence that match-description lineups, compared to similarity-based lineups, yield superior correct identification rates for culprit-present lineups with no difference in false identifications from culprit-absent lineups (Wells et al., 1993). There are, however, unresolved issues regarding the use of a match-description strategy for foil selection. As acknowledged by advocates of the match-description strategy (Wells, 1993; Wells et al., 1998), the description provided by witnesses may be inaccurate, or too vague to provide sufficient guidance for foil selection (Meissner, Sporer, & Schooler, 2007). In such cases, the use of a match-description strategy may well produce an unfair lineup in which the suspect stands out as different from other lineup members. These problems are not necessarily fatal for the match-description strategy, but they do highlight the point that, alone, it often does not provide a sufficient basis for foil selection.

In sum, there is clear evidence that the selection of foils can substantially influence responses to a lineup, and that the use of foils that are either implausible or too similar to the suspect is problematic. However, progress towards an effective solution to this problem requires a greater understanding of the effects that variables such as functional size and foil similarity have on identification decisions. One promising line of research, which would aid the systematic investigation of these variables, involves the development of reliable methods for quantification of facial similarity (Tredoux, 2002).

**Lineup presentation**

Here, we examine some of the key issues in presenting the lineup, including the medium of presentation (e.g. live lineup vs. photoarray), the lineup administrator, the interrelationships between the lineup and other formats such as mugshots and showups, and the interactions between the lineup administrator and the witness prior to and during the conduct of the lineup.
Lineup presentation medium

It is perhaps surprising that, given the overall direction of this paper, we discuss an issue on which we believe no clear and empirically supported direction can be given. We do so, however, because the available evidence is almost certainly at odds with commonly held views. Although, we are not aware of any data to support the following statement, we suspect that it would be a common view amongst forensic practitioners that a live lineup will produce more accurate identification responding than any kind of photoarray. A live lineup provides an impression of height and build, allows the witness to see lineup members from different angles and to see them stationary and moving. Surely the availability of such cues will inform the identification response in ways that the (static) photoarray cannot.

While it is possible that this intuitively appealing perspective may eventually be validated by adequately powered studies incorporating a variety of stimulus materials, there is not, at present, a sound body of supportive empirical evidence. Indeed, the available evidence points to an absence of meaningful differences in identification performance between live lineups and photoarrays. Most of the available evidence was reviewed by Cutler, Berman, Penrod, and Fisher (1994). It is worth noting that a number of the individual published studies on this issue are characterized by important limitations that make interpretation extremely difficult, thereby challenging any reliance on the results of any single study. The limitations include small sample sizes (Cutler & Fisher, 1990; Egan, Pittner, & Goldstein, 1977; Shepherd, Ellis, & Davies, 1982), failure to include perpetrator-absent lineup conditions (Egan et al., 1977; Shepherd et al., 1982), and various methodological confounds (Cutler & Fisher, 1990). Examination of results aggregated across studies, as reported in Cutler et al. (1994), suggests that (1) false-alarm rates do not differ significantly for photo and live lineups, and (2) hit rates are slightly (but not meaningfully) higher for live than photo lineups. One more recent empirical paper providing such contrasts is Morgan et al. (2004), though this paper has some significant methodological limitations. Nevertheless, their data are also generally in line with the above conclusions: specifically, they found no meaningful differences in hit rates, and a possible small advantage for the photolineup with respect to false alarms.

Assuming the conclusions based on the above studies withstand future replications, why might it be that the live lineup does not prove superior to a photoarray? Perhaps the encoding conditions experienced by witnesses to crimes are generally so unfavourable that witnesses, on average, are unlikely to 'capture' the cues that they might be able to exploit given a live lineup. Or perhaps the degree of match between encoding and test stimuli is typically sufficiently low that an accurate identification response is dependent on the individual witness encoding a memorial image of sufficient quality that compensates for any of the supposed limitations of photographic stimuli. Regardless of the reason, the empirical data available at present suggest that there are many variables in lineup conduct that are more important in shaping identification performance than the lineup presentation medium.

The lineup administrator

Who conducts the lineup is a key issue. Should it be the investigating police officers or an independent administrator with no knowledge of the case? Or perhaps the lineup should be conducted on a computer? And, why should it matter?

It matters because the identification test involves some degree of interaction between the witness and the administrator, which of course means it offers
opportunities for social influence. Consider the following. Most witnesses asked to view a lineup are likely to infer that the lineup contains a police suspect, with their task being to assist the police case by confirming that the suspect is the culprit. This belief on the part of a witness almost certainly biases them towards making a lineup selection rather than a rejection. Moreover, should the lineup administrator's interactions with the witness - whether they be calculated or inadvertent, explicit or subtle - strengthen the latter's belief that the police 'have' the culprit (e.g. 'We'd like you to come down and try to pick out the culprit' or 'We reckon we've just about cracked this case and would like you to take a look at a lineup').), any tendency to want to choose from the lineup is likely to be enhanced. Lineup administrator behaviour that biases a witness towards a positive identification decision obviously can have devastating consequences if the outcome is the misidentification of an innocent suspect.

Further and potentially more damaging opportunities for an administrator to influence a witness arise when the administrator actually presents the lineup to the witness. If the lineup administrator happens to be the investigating officer, or someone who has communicated with the investigator about the suspect, there is obvious potential to cue the witness as to which lineup member is the suspect, and which are merely innocent foils. (See, for example, Brewer & Wells, 2009, for a series of examples of the types of interactions between administrator and witness that have the potential to cue the witness in this way). Again, these cues as to the identity of the suspect or the foils might be provided quite deliberately by the administrator or might arise quite innocently as the interaction unfolds.

Clearly, given the adverse consequences of a misidentification, some form of protection against the influence of the lineup administrator is required. Two main steps have been recommended. The first involves double-blind lineup administration, an approach advocated by Wells (1988) and long accepted as a way of reducing or eliminating demand effects in human testing procedures. The second involves providing witnesses with unbiased (or non-leading) instructions which are designed to counter any systematic bias towards making a positive identification decision.

Double-blind lineup administration
Double-blind administration (cf. single-blind) simply means that the lineup administrator has no knowledge of which lineup member is suspected of being the perpetrator and which lineup members are innocent foils. In other words, the investigating officer has no role to play in the presentation of the lineup, thereby (hopefully) removing any opportunity for the administrator to shape the witness's decision. Achieving this independence between the investigator and the lineup administrator obviously requires the establishment of rigid organizational protocols for testing, which could of course include computerized lineup presentation and recording of witness responses (as used in some police jurisdictions).

Despite any practical difficulties associated with double-blind lineup administration, empirical research suggests that the effort should be worthwhile. In addition to the substantial literatures in mainstream psychology demonstrating the power of the expectations of experimenters or other people conducting testing, several studies have contrasted the effects of double- and single-blind lineup administration (e.g. Greathouse & Kovera, 2008; Haw & Fisher, 2004; Phillips, McAuliff, Kovera, & Cutler, 1999). Although, the findings from these studies are not consistent, all have detected effects indicative of lineup administrators influencing witnesses' decisions.
Phillips et al. (1999) found higher rates of false identifications of innocent suspects under single-blind (cf. double-blind) conditions if an independent observer was watching the administration (perhaps, as they hypothesized, 'pressuring' the administrator to secure a positive identification). Haw and Fisher (2004) found similar effects, at least when the administrator had close contact with the witness. Finally, Greathouse and Kovera (2008) found effects for both culprit-absent and culprit-present lineup decisions under conditions which lowered the witness's decision criteria (i.e. induced guessing), with the diagnosticity of identification decisions higher under double-blind administration. Moreover, Greathouse and Kovera also found that, while witnesses and lineup administrators seemed not to be aware of any behavioural differences in administrators under the two conditions, independent observers noted that single-blind administrators were more likely to behave in ways that might promote choosing (e.g. telling the witness to look carefully or to take another look when they didn't pick, or indicating that they knew who the suspect was).

Notably, such effects have been detected in the eyewitness laboratory using student lineup administrators who, arguably, (a) have less motivation than police investigators to secure a pick of the suspect, and (b) less experience in the ways of doing so. Together, the findings - especially when coupled with the broad psychological literature on demand effects - suggest that, regardless of any practical difficulties associated with its implementation, double-blind lineup administration represents a valuable protective procedure.

Unbiased lineup instructions

Unbiased, or warning, instructions clearly communicate to (or warn) the witness, prior to viewing the lineup, that the culprit may or may not be in the lineup. In contrast, biased, or no warning, instructions fail to include the latter warning. Of course, instructions may be even more leading or biased than no warning instructions if, for example, the lineup administrator instructs the witness in a way that somehow implies they will be able to pick the culprit if their memory is sound or that response options such as not sure or don't know are undesirable. In contrast, unbiased instructions clearly signal to the witness that a positive identification decision is not the only response option and that it is quite appropriate to respond with options such as not there or not sure.

Biased (or no warning) instructions increase choosing. Given that, on average, the best match to the witness's memory will be the culprit, it is not surprising that, for culprit-present lineups, biased instructions sometimes result in a higher proportion of culprit selections as well as more foil selections (Brewer & Wells, 2006; Clark, 2005). Importantly, the proportion of witnesses who make positive decisions when presented with culprit-absent lineups is significantly reduced by the warning re the possible absence of the culprit (Brewer & Wells, 2006; Malpass & Devine, 1981; Steblay, 1997). In sum, while some form of biased instructions may well increase the likelihood that the police suspect is identified, the costs are high and emphasize the importance of a clear warning regarding the possible absence of the culprit.

These patterns not only characterize adult witnesses but also child witnesses (e.g. Keast, Brewer, & Wells, 2007). Research clearly shows that children have a propensity to make a positive decision, with this consistently reflected in higher proportions of positive identifications from culprit-absent lineups than is found for adults (e.g. Keast et al., 2007; Parker & Carranza, 1989; Parker & Ryan, 1993; Pozzulo & Lindsay, 1998).
Various approaches to controlling or overcoming this tendency have been examined, including providing additional response options that allow children still to pick a stimulus when rejecting a lineup (Beal, Schmitt, & Dekle, 1995; Davies, Tarrant, & Flin, 1989; Zajac & Karageorge, 2009), practice lineups (Davies, Stevenson-Robb, & Flin, 1988; Parker & Ryan, 1993), different lineup formats such as the elimination lineup (Pozzulo & Lindsay, 1999), and accuracy motivation instructions (Brewer, Keast, & Sauer, 2009; Pozzulo & Lindsay, 1997). While each of these interventions has sometimes shown promise, consistent and reliable empirical support has not yet been found for any of these interventions. This makes the clear provision of unbiased instructions which emphasize the appropriateness of not having to make a pick from a lineup absolutely critical for child witnesses.

The relationship between the lineup and other identification tests

Lineups probably represent the most commonly held view of what an eyewitness identification test comprises. However, two different identification procedures are also widely used by police: Pre-identification mugshots and showups. Each of these tests differs in potentially important ways from an identification lineup.

Pre-identification mugshots

In order to narrow the search for a suspect, police investigators will often ask witnesses to search through a series of mugshots. The chances of a witness correctly identifying the culprit decrease as the number of faces viewed prior to encountering the culprit’s face increases (Laughery, Alexander, & Lane, 1971; Lindsay, Nosworthy, Martin, & Martynuck, 1994; McAllister, Stewart, & Loveland, 2003). There is also some evidence that it is the number of similar, rather than dissimilar, faces viewed that produces this effect (Lindsay et al., 1994).

Mugshot procedures can also be considered in terms of the effects they might have on a subsequent identification decision, should the witness later be asked to view a lineup. Several experiments have found no evidence that mugshot searches affect subsequent identification decisions (Cutler, Penrod, & Martens, 1987; Cutler, Penrod, O’Rourke, & Martens, 1986; Dysart, Lindsay, Hammond, & Dupuis, 2001), with one important exception. If an innocent face viewed during an initial mugshot search also appears in a later lineup, that face is more likely to be chosen, presumably because it appears more familiar to the witness (Brown, Deffenbacher, & Sturgill, 1977; Deffenbacher, Carr, & Leu, 1981; Perfect & Harris, 2003). This effect is even stronger if the face was actually chosen by the witness during the initial mugshot search (Deffenbacher, Bornstein, & Penrod, 2006; Dysart et al., 2001). The clear implication of these findings is that a witness who has been exposed to a suspect’s face during a mugshot search should not then view a lineup containing that suspect.

Showups

Showups involve the presentation of a single suspect, rather than a lineup, to a witness. This method is used quite often by police, with several field studies reporting that showups comprise between 30 and 77% of identification tests conducted in real cases (see Dysart & Lindsay, 2007, for a review).
Showups are generally frowned upon by eyewitness experts as being too suggestive (e.g. Kassin, Tubb, Hosch, & Memon, 2001). However, in a meta-analysis comparing showups with lineups, Steblay, Dysart, Fulero, and Lindsay (2003) found some interesting and counter-intuitive results. First, witnesses were approximately twice as likely to make a positive identification from a lineup as they were from a showup. Second, the correct identification rates obtained from culprit-present trials with the two procedures were virtually identical. Third, the correct rejection rate was much higher for culprit-absent showups than lineups. Fourth, across culprit-present and culprit-absent lineups combined, accuracy was significantly greater for showups than lineups. Together, these results appear consistent with the idea that showups promote the use of absolute rather than relative judgments (Gonzalez et al., 1995). However, positive identifications from culprit-absent lineups include false identifications of innocent suspects and known-foil identifications, whereas all positive identifications from culprit-absent showups are false identifications. When the false identification rate for lineups was calculated as the rate of positive responses to culprit-absent lineups divided by the number of lineup members, this equated to approximately half the rate of false identifications found using showups.

In sum, compared to lineups, showups produce lower choosing rates and comparable correct identification rates. However, the absence of foils in the showup procedure results in a higher rate of false identifications of innocent suspects than is found for lineups.

**Presentation format: Simultaneous versus sequential lineups**

The two best-known forms of lineup presentation are simultaneous presentation, whereby a witness is shown all lineup members at once, and sequential presentation, whereby a witness is shown one lineup member at a time and asked to indicate whether each is the culprit or not. The issue of whether sequential lineup presentation is superior to simultaneous presentation has perhaps stimulated more debate than any other among eyewitness researchers (e.g. Lindsay, Mansour, Beaudry, Leach, & Bertrand, 2009; Malpass, Tredoux, & McQuiston-Surrett, 2009). Sequential lineups produce lower choosing rates, or rates of positive decisions, than simultaneous lineups. Lower rates of positive decisions translate into fewer false identifications from culprit-absent lineups and fewer correct identifications from culprit-present lineups (e.g. Carlson, Gronlund, & Clark, 2008; Lindsay & Wells, 1985; Steblay et al., 2001). The central claim in favour of sequential lineups is that they reduce false identifications to a greater extent than they reduce correct identifications, resulting in a net overall gain in accuracy (Lindsay & Wells, 1985; Steblay et al., 2001). Proponents of the sequential lineup have claimed that the accuracy advantage arises because sequential presentation encourages witnesses to make absolute (i.e. does this face match my memory of the offender?) rather than relative (i.e. which face is the best match to my memory of the offender?) judgments (e.g. Lindsay et al., 2009; Lindsay & Wells, 1985). In contrast, it has been argued that sequential lineups simply produce a conservative criterion shift (e.g. Flowe & Ebbesen, 2007), with witnesses requiring more evidence for making a positive decision. Moreover, it has been argued that the sequential advantage is only found when biased lineups (i.e. one lineup member stands out from the others) are used (suggesting that, at worst, the sequential lineup may be a useful 'protective' procedure), and varies depending on where the suspect appears in the sequence (Carlson et al., 2008).
Despite the enormous amount of recent debate surrounding this topic, we are perhaps at odds with many researchers in this field in that we do not consider this to be a central issue. As acknowledged by Lindsay et al. (2009), neither simultaneous nor sequential lineup presentation is likely to produce acceptable accuracy rates. Thus, while lineup presentation mode is worthy of close investigation by researchers, our view is that considerable research attention should be directed towards the consideration and development of novel, more effective procedures, rather than focusing so much on the evaluation of competing existing ones.

Recording the witness's decision
It seems trite to suggest that, when an eyewitness reports their decision after viewing the lineup, this decision should be faithfully recorded. After all, it is a key component of the evidence and it obviously will be recorded somehow. However, it is important that witness responses are recorded with precision as retrospective accounts of events often deviate from what actually occurred. Moreover, there are accompanying witness behaviours that are potentially informative about the likely accuracy of the witness's decision that are subject to distortion by events that follow the identification test and, hence, should be recorded promptly after that test. Ideally, there are three things that should be recorded at the time the identification response is made: (1) the exact response made by the witness, (2) the witness's confidence in their decision, (3) how long the decision took (i.e. decision latency).

The identification response
Some witnesses might say quite decisively 'that's him' or 'she's not there'. Indeed, if the lineup involves a photoarray conducted on a computer, the various possible response options can be specified precisely. However, in the absence of the response specificity that a computerized presentation can demand, a witness might also say things like 'it could be either two or six, maybe two' or 'I'm not sure but number two looks close to him'. Both of these options could conceivably enter the record of evidence as a choice of lineup member two, especially if two is the police suspect - but they are quite different responses to 'it's number two' and clearly say something quite different about the degree of match between the witness's memory and lineup member two. So, obvious as it may seem, recording witness decisions precisely is critical if identification test evidence is to provide an accurate reflection of the witness's decision.

Identification confidence
Expressions of confidence in the eyewitness identification decision often accompany the decision. These judgments might be volunteered by the witness at the time of making the decision or solicited by the police investigator. It is well documented that confidence in the identification influences other people's (e.g. police, lawyers, judges, and mock-jurors) judgments of the likely reliability of the eyewitness identification evidence (e.g. Bradfield & Wells, 2000; Brewer & Burke, 2002; Cutler, Penrod, & Stuve, 1988; Lindsay, Wells, & O'Connor, 1989; Potter & Brewer, 1999). Despite repeated assertions from psychologists that confidence judgments provide no useful basis for evaluating the identification decision, the current state of knowledge suggests a different, and more complicated story.
If - and only if - identification confidence is assessed immediately after the identification, and prior to the witness receiving any feedback about their decision, the evidence suggests the following conclusions (see for example, Brewer, 2006; Brewer & Weber, 2008; Brewer & Wells, 2006; Juslin, Olsson, & Winman, 1996; Keast, Brewer, & Wells, 2007; Lindsay, Nilsen, & Read, 2000; Lindsay, Read, & Sharma, 1998; Sauerland & Sporer, 2009; Sporer, Penrod, Read, & Cutler, 1995). First, high confidence in the identification decision does not ensure that the decision was accurate. All of the above data sets illustrate that extremely confident identifications are sometimes inaccurate. Second, despite there usually being some degree of overconfidence, confidence-accuracy calibration research shows that extremely confident identifications are associated with relatively high degrees of accuracy whereas unconfident decisions are not. In other words, a very confident decision provides a strong pointer to the police that their suspect is a viable one and at least worthy of continued investigation. In contrast, an unconfident decision strongly suggests that they need to consider other hypotheses about the culprit’s identity. Note, however, that this conclusion applies only to the evaluation of positive decisions (i.e. lineup choices). Confidence appears to provide no useful information about the likely accuracy of negative decisions (i.e. lineup rejections), despite the fact that lineup rejections point to a mismatch between the suspect and the witness's memory (Wells & Olson, 2002). Furthermore, it does not extend to decisions made by children aged around 10–12 years or so (and, almost certainly, not to younger children), for whom identification confidence offers no useful guide to accuracy.

In sum, identification confidence can be a very useful piece of information for police investigators in terms of directing, or redirecting, their focus to particular suspects. But the confidence assessment should be an independent and immediate witness assessment.

**Identification latency**

Some witnesses make rapid decisions when confronted with a lineup; others deliberate for lengthy periods. The evidence does not allow us to specify some magical decision latency which effectively diagnoses identification accuracy. But there are two reasons why we believe it is important to consider this variable here. First, there has been some suggestion in the literature that it may be possible to diagnose accuracy using decision latency. For example, Dunning and Perretta (2002) reported that decisions made within 10–12 s were nearly always correct. Subsequent research refutes that conclusion, clearly showing that the decision latency that best discriminates accurate from inaccurate decisions is highly variable (Brewer, Caon, Todd, & Weber, 2006; Sauer, Brewer, & Wells, 2008; Weber, Brewer, Wells, Semmler, & Keast, 2004). Moreover, in the absence of an identification test format (viz. computerized presentation) that standardizes parameters such as presentation mode, stimulus size and arrangement, instructions, and so on, this is always likely to be the case.

Second, identification decision latency is reliably associated with decision accuracy for positive (though not negative) decisions, with accurate decisions faster (on average) than inaccurate (Brewer & Weber, 2008; Brewer et al., 2006; Brewer, Gorden, & Bond, 2000; Sauerland & Sporer, 2009; Sporer, 1992, 1993, 1994; Weber et al., 2004). It is, therefore, potentially extremely useful to record this characteristic of a decision at the time the decision is made, so that at later stages of an investigation and/or trial it is possible to provide an accurate characterization of the decision. Not only should police have some reservations about their suspect if a witness took an extremely long time to
make a decision, but the courts should also have similar reservations about the identification decision of a witness who testifies that s/he is absolutely certain the defendant is the culprit and yet the record of the decision process suggests the witness had extreme difficulty reaching a decision at the lineup. In contrast, any reservations about a witness’s identification decision should be allayed somewhat when the witness made an extremely rapid identification and expressed a very high degree of confidence at the time of the decision (Weber et al., 2004).

Feedback to the witness
There is now a substantial body of empirical research demonstrating that significant influences on identification test evidence can occur even after the witness has made a decision. Specifically, following an identification decision, witnesses will often be the recipients of feedback. Indeed, if they behave anything like laboratory participants they are likely to actively seek feedback that might validate their decision. This feedback might come from a lineup administrator (e.g. ‘Good, you identified the suspect’), the investigating officer (e.g. ‘You would testify to that in court, right?’ or another witness (e.g. ‘Which one did you pick? ... yeh, I thought that was him too.’). The feedback might be non-verbal or verbal in nature (e.g. gestures or expressions), confirming (as in the preceding examples), or disconfirming (e.g. ‘That’s ok, not everyone gets a good enough look to guarantee them a strong memory of the guy.’). Here, we will consider the impact of feedback on (1) subsequent identification performance, should a witness be asked to attend a second identification test, and (2) identification confidence and other judgments about both the witnessing conditions and the identification test.

Effects of feedback on subsequent identification performance
Occasionally, witnesses may be asked to view more than one lineup during an investigation. For example, if a crime was committed by more than one culprit, investigators might be expected to construct an appropriate lineup for each suspect. Multiple lineups might also be used in single-culprit crimes. For example, if there are multiple suspects for a single-culprit crime, the police might construct a separate lineup for each suspect and ask the witness to view each lineup in turn (presumably stopping when a suspect is chosen, or all lineups have been viewed). Alternatively, the police may arrest one suspect and have the witness view a lineup containing that person. Later, new evidence may emerge exonerating the first suspect and implicating another person. If so, the witness may then be asked to view a lineup containing the second suspect.

Little is known about the effects of postidentification feedback following one identification test on responses to a second test. However, recent data from our laboratory provide some clear evidence that disconfirming feedback following an initial lineup response is very likely to impair accuracy on a second test. This pattern of results has been found when a witness chooses someone from an initial lineup and is then asked to view a second lineup for the same culprit (Palmer, Brewer, & Weber, 2009a), and when a witness rejects an initial lineup for one culprit and is then asked to view a second lineup for a different culprit (Palmer, Brewer, & Weber, 2009b). Thus, based on the available data, our recommendation would be to avoid giving disconfirming feedback following an initial identification decision if there is a chance that the witness may subsequently be asked to view another lineup.
Effects of feedback on confidence and other judgments

While the availability (and seeking) of such feedback seems perfectly natural, we now know that it can exert a dramatic effect on the identification test evidence. Previously, we outlined the value of confidence judgments obtained immediately after the identification. The usefulness of such judgments reflects the fact that they provide a 'purier' measure of memory strength than do judgments that have been influenced by postidentification interactions with investigating police or other witnesses. Indeed, even very brief and apparently benign postidentification interactions have been shown to exert a powerful influence on subsequent confidence judgments.

Confirming feedback often dramatically inflates witness confidence, whereas disconfirming feedback produces confidence deflation (Bradfield, Wells, & Olson, 2002; Hafstad, Memon, & Logie, 2004; Luus & Wells, 1994; Wells & Bradfield, 1998, 1999; Wells, Olson, & Charman, 2003). This pattern has been detected for positive identifications and for lineup rejections from both culprit-present and culprit-absent lineups (Semmler, Brewer, & Wells, 2004). It can be produced by feedback from a 'live' administrator, a computer or a co-witness (Luus & Wells, 1994; Semmler et al., 2004). The effect has even been demonstrated when a lineup administrator did not provide verbal feedback, but knew the suspect's identity, thereby illustrating how non-verbal cues can produce the effect (Garrioch & Brimacombe, 2001).

Why is this a problem? Confidence inflation produced by confirming feedback clearly can turn a hesitant witness - who might say he was 75% confident at the time of making the identification when his memory was relatively strong - into an extremely confident one - expressing say 90-100% confidence - at some later date when his memory is likely to be weaker. And as indicated earlier, highly confident testimony constitutes very persuasive evidence. Similarly, the confidence-deflating effects of disconfirming feedback following a witness picking a known-innocent lineup foil is also a problem. By reducing the witness's confidence, and hence believability, in the choice that was made, the effect is to undermine the value of identification evidence that is known to be of diagnostic value (Wells & Olson, 2002).

The consistent and robust findings on the malleability of confidence reinforce our earlier point about confidence being diagnostic of accuracy, only if recorded at the time of the identification. In contrast, confidence reports solicited in the courtroom should be considered uninformative and quite likely highly misleading with respect to the reliability of the witness's memory.

Postidentification feedback also has an effect on other witness judgments, specifically those relating to perceptions of the witnessing and identification test experiences. Witnesses' perceptions of the witnessing conditions include things like how good a view they had of the culprit and how closely they were paying attention, whereas perceptions of the identification test include things like whether the decision was easy to make or whether one of the lineup members just 'popped out' as the culprit. Researchers have shown that, compared to no feedback controls, confirming feedback leads to witnesses reporting having experienced better quality encoding conditions, paying closer attention, and making easier and more rapid identifications (Wells & Bradfield, 1998, 1999), although these effects appeared to be confined to responses to subjective (cf. objective) questions (Douglass, Brewer, & Semmler, 2009; Douglass & Steblay, 2006). Like witness confidence, such witness reports have the potential to influence jurors' assessments of the overall quality of the evidence, reinforcing the importance of also recording such perceptions and judgments at the time of the identification.
**Summarising what works**

In sum, what should lineup practitioners focus on given the current state of knowledge? What are the procedures most likely to ensure maximum diagnosticity for eyewitness identification test evidence? The following points capture the main conclusions to emerge from the preceding discussion:

1. Lineups should contain only one suspect. Foils should capture the key features of the witness’s description and be plausible matches for, but not too similar in appearance to, the suspect.
2. Do not be preoccupied with the lineup presentation medium as there are many other factors that seem to be far more important.
3. Double-blind lineup administration should be preferred.
4. Witnesses who have previously seen a suspect’s face in a mugshot search or showup should not subsequently view a lineup for that suspect.
5. Ensure unbiased instructions providing clear warnings regarding possible culprit absence.
6. Immediately after the identification decision, ensure there is an (independent) record of the witness’s exact identification response, the confidence in that decision, the witness’s decision latency, and other perceptions that witnesses had of the encoding and identification experiences.
7. Avoid giving disconfirming feedback to witnesses following an initial identification decision if there is a chance that the witness may subsequently be asked to view another lineup.
8. Discount courtroom expressions of identification confidence and other witness perceptions of encoding and the lineup.

At this point it is important to reiterate a point we made at the start. The available evidence clearly suggests that, even if all of the suggested guidelines are diligently observed, eyewitnesses will still make inaccurate decisions when examining lineups, including the selection of innocent suspects and lineup foils and the rejection of lineups containing the culprit. Moreover, while we are hopeful that future work will gradually help us refine these guidelines, we believe that the traditional identification test formats that we have been discussing in this chapter will remain limited in their capacity to discriminate the guilty from the innocent.

**Contributions of research to the operation of the justice system**

Eyewitness memory researchers have identified numerous ways to improve the collection and interpretation of eyewitness evidence. Importantly, at least some of these advances in understanding have been translated into policy change. For example, the Police and Criminal Evidence act (UK Home Office, 2008) includes instructions to warn witnesses prior to viewing a lineup that the perpetrator may or may not be present in the lineup. The National Institute of Justice Technical Working Group for Eyewitness Evidence (1999) guidelines also include, for example, instructions for lineup administrators to use single-suspect lineups, and to record identification confidence before witnesses have access to feedback as to whether the suspect was identified. Each of these procedures has been shown through research to enhance the reliability of eyewitness identification test evidence and, thus, has likely contributed to reducing mistaken identification and increasing the effectiveness of police investigations.
However, despite these and other advances, the legal system is not yet clamouring for psychologists to tell them how best to conduct identification tests. The capacity for research to influence policy would benefit from a greater level of interaction between psychologists and legal professionals. The openness of law enforcers to the influence of research findings is likely to increase with greater familiarity with those findings, and the credibility of researchers in presenting the findings is likely to increase with greater appreciation of the workings and culture of the law enforcement and legal professions.

**Future trends**

As outlined above, implementation of the various procedures suggested by empirical research by no means guarantees reliable eyewitness identifications. Regardless of the testing conditions, rates of misidentifications and rejections of lineups containing the culprit remain higher than would be desirable. To some degree, this state of affairs is inevitable as the justice system obviously has no control over the encoding conditions experienced by witnesses and, hence, the initial quality of their memories. But leaving aside the limitations on identification test performance imposed by impoverished memorial images, our hope is that future research will find novel and effective ways of enhancing our ability to tap eyewitnesses’ memories.

We suspect that any substantial advances from here on may well involve radical departures from previous methods. Consider some recent comments from Wells *et al.* (2006):

It could be argued that research has been profoundly conservative in its approach to the eyewitness-identification problem. Specifically, researchers have tended to operate within the confines of the traditional lineup, in which a suspect is placed among fillers and the eyewitness makes a verbal identification. But what if the lineup had never existed and the legal system turned to psychology to determine how information could be extracted from eyewitnesses’ memories? ... Operating from scratch, it seems likely that modern psychology would have developed radically different ideas. For instance, brain-activity measures, eye-movements, rapid displays of faces, reaction times, and other methods for studying memory might have been developed instead of the traditional lineup. Once we step outside the confines of the traditional lineup, it is possible to imagine a future science of eyewitness evidence that is radically different from the methods used today (pp. 68–69).

The approaches taken in some recent research align with this general message. For example, R.C.L. Lindsay and colleagues have experimented with multiple lineups, testing witnesses across several lineups including faces, clothing, voices, and bodies (Pryke, Lindsay, Dysart, & Dupuis, 2004). While false identification rates are addressed by such procedures, hit rates are problematic. A radically different approach currently being pursued in our laboratory (a) asks witnesses to indicate their confidence that each lineup member is the culprit instead of asking them to either make a specific choice or reject the lineup, and (b) tests how effectively algorithms developed around the patterns of those confidence judgments identify whether a lineup member is/is not the culprit. Early results are promising, suggesting that this approach more effectively identifies whether a lineup member is/is not the culprit than does the traditional simultaneous lineup decision (e.g. Sauer, Brewer, & Weber, 2008). Thus, one potentially fruitful way forward from here will be to search vigorously for completely new alternatives for tapping eyewitness memory that, nevertheless, draw upon known principles of memory functioning.
Acknowledgements

This research was supported by grant DP0556876 from the Australian Research Council and a Flinders University Research Grant.

References


Received 3 December 2008; revised version received 5 January 2009