

2.29 Eyewitness Identification

Melissa F. Colloff

University of Birmingham, Birmingham, United Kingdom

Laura Mickes

University of Bristol, Bristol, United Kingdom

Scott D. Gronlund

University of Oklahoma, Norman, OK, United

Abstract

Eyewitnesses commonly provide evidence during criminal cases, but are commonly perceived as error-prone. We present results that challenge this notion. The core of this challenge arises from viewing research from the joint perspectives of discriminability (the ability to distinguish innocent from guilty suspects) and reliability (the likelihood the person identified is actually the perpetrator). We also describe an organizational framework that classifies variables according to the different stages of memory as they map onto the different stages of the crime/criminal justice process, and review how these variables affect discriminability and reliability. Although particular variables impair witness discriminability, witnesses still are often reliable, because identifications made with high confidence on the initial memory test are typically highly accurate. The discriminability/reliability perspective and our organizational framework have implications for how eyewitness identification evidence should be used by the criminal justice system.

Keywords: Confidence accuracy characteristic analysis; Confidence–accuracy relationship; Diagnosticity ratio; Eyewitness discriminability; Estimator variables; Eyewitness identification; Eyewitness memory; Lineup; Receiver operating characteristic analysis; Eyewitness reliability; Showup; System variables

Acknowledgments

This work was partially supported by Economic and Social Research Council grants to Laura Mickes [ES/L012642/1] and Melissa Colloff [ES/T01475X/1]. We thank Kayolan Ganev for his valuable assistance on an earlier version of this chapter.

Key Points

- Eyewitness identification evidence plays a pivotal role in how a case proceeds through the criminal justice system.
- Witnesses are commonly perceived as error-prone, but a different view transpires when two elements of witness performance are considered: discriminability (the ability to distinguish innocent from guilty suspects) and reliability (the likelihood the person identified is the perpetrator).
- We describe an organizational framework that classifies variables that may affect witness memory performance (e.g., witnessing factors, like distance). The framework organises the variables according to the different stages of memory as they map onto the different stages of the crime/criminal justice process.
- Although particular variables impair witness discriminability, witnesses still are often reliable, because identifications made with high confidence on the initial memory test are often highly accurate.
- The discriminability/reliability perspective and organizational framework has implications for how eyewitness identification evidence should be used by the criminal justice system.

2.29.1 Introduction

Eyewitnesses commonly provide evidence in criminal cases, but have a bad rap for being error-prone. Although eyewitnesses have difficulty withstanding suggestions and feedback, their memories also have been criticized due to an inability to use confidence levels to reflect the likely accuracy of their identification. After all, faulty eyewitness evidence has played a role in 70% of the wrongful imprisonment of the innocent (Innocence Project, 2019). Is the notorious reputation merited? Can eyewitness memory never be trusted? This chapter presents results that challenge that notion.

2.29.1.1 An Eyewitness Task

One way police investigators glean information from an eyewitness is to administer an identification procedure, which is essentially a recognition memory test of the eyewitness's memory for the perpetrator. Identification procedures include lineups (photos, videos, or live), showups, field views, and mugbooks (National Research Council, 2014). Live lineups are perhaps what many people imagine when thinking about police lineups—several individuals stood in a row, with the witness nervously walking along the line, and then pointing at the person they believe to be the perpetrator. Although live lineups are still used in some countries (e.g., South Africa), they have been superseded in many places with photo or video lineups (e.g., US, UK, Australia), in which the witness is presented with images of the lineup members. In a showup, there are not multiple lineup members; the witness is shown the police suspect and asked if the suspect is the person who committed the crime. Showups are typically used when a suspect who fits the witness's description has been found in the vicinity of a recent crime. Field views and mugbooks are used when the police do not have a suspect. In a field view, the police invite a witness to view many people in the context where the perpetrator might likely appear,

such as a busy shopping area that is a crime hotspot. In a mugbook, the witness reviews a collection of photos, such as local offenders, or a student yearbook if the perpetrator might be from a specific school.

Eyewitness evidence plays a pivotal role in how a case proceeds through the criminal justice system. A witness identification of a police suspect is considered compelling evidence of guilt, and suspects who have been identified are more likely to be charged and found guilty than those who have not (e.g., Devlin, 1976). In many cases, the quality of the witness memory evidence is vital, because other types of evidence, such as DNA, physical injury, or CCTV, is unavailable (e.g., Kelly et al., 2005).

Because the correct collection and interpretation of witness evidence is critical, scientists study eyewitness memory in controlled laboratory conditions. Participants are typically asked to watch a mock-crime (often a video) and then, after a delay, are presented with an identification procedure and asked if they can identify the perpetrator from the video. Sometimes the perpetrator is present (target-present), and sometimes absent (target-absent). Target-present procedures represent the real-world situation in which the police suspect is guilty, and target-absent lineups represent the situation in which the police suspect is innocent. The benefit of a laboratory procedure like this is that variables can be carefully controlled or manipulated and, unlike in real crimes, researchers know who is the perpetrator. The variables that affect eyewitness accuracy are often measured using lineups (photos or videos) or showups, and that is what we focus on in this chapter.

2.29.1.2 Lineups

A standard police lineup contains one suspect (who is either guilty or innocent) and multiple fillers (people who are known to be innocent). Fig. 1 shows a schematic of a six-person simultaneous photo lineup. The images of each lineup member (photos or videos) are either simultaneously or sequentially presented, depending on the policy adopted by a particular country or by a particular jurisdiction within a country. The number of lineup members typically used also differs around the world (e.g., Fitzgerald et al., 2021).

There are three possible choices that eyewitnesses can make when faced with a lineup. They can identify (1) the suspect, (2) one of the fillers, or (3) no one. All possible outcomes are in Table 1. If the target (i.e., perpetrator) is present in the lineup, and the eyewitness identifies them, then that is a correct identification (also known as a hit). If the target is not present in the lineup, and the eyewitness identifies the innocent suspect, then that is a false identification (also known as a false alarm). These decisions are in red-bolded font because these are the decisions that matter, and thus the focus is typically placed on the identification of the guilty and innocent suspects. That is because fillers are known to be innocent, and although picking one is an error, there is no risk of the filler being wrongfully investigated or convicted. If the eyewitness does not identify a lineup member, and if the target is in the lineup, then that is a miss; and if the target is not present in the lineup, then that is a correct rejection. In a laboratory study, the correct ID rate is the proportion of guilty suspects identified from the target-present lineups. The false ID rate is the proportion of innocent suspects identified from the target-absent lineups. Correct ID rates and false ID rates are then used to assess the discriminability of a lineup procedure (see Discriminability).

2.29.1.3 Showups

There are only two possible choices that eyewitnesses can make when faced with a showup. They can (1) identify the suspect, or (2) not. If the suspect is the perpetrator, and the eyewitness identifies them, then that is a correct identification (a hit). If the suspect is not the perpetrator, and the eyewitness identifies the innocent suspect, then that is a false identification (a false alarm). Correct and False ID rates are calculated in the same way as for lineups.

2.29.1.4 System Versus Estimator Variables

Examining lineups and showups, researchers have dissected the variables that affect eyewitness accuracy and studied each constituent part. Those parts often have then been grouped into whether they can be used to make decisions to improve the legal system (e.g., how to conduct a lineup, or what instructions to read to an eyewitness) or whether they involve variables that affect the accuracy of an identification (e.g., a cross-race identification, length of the retention interval). These variables typically have been categorized as system and estimator variables (Wells, 1978), respectively. A system variable is under the control of the legal system and an estimator variable is not. That distinction, however, only applies to policymakers who need to make decisions about how eyewitness evidence should be gathered by the police, because how a lineup is conducted can be specified, but whether the eyewitness and the perpetrator are the same race cannot. However, there is no such system/estimator distinction when it comes to the court of law. That is, for judges and jurors who need to determine the likely culpability of a defendant based on the identification of an eyewitness, there is no control over any variable whatsoever. Take, for example, lineup presentation. Whether to present an eyewitness with a simultaneous lineup (where all the faces are presented together) or sequential

lineup (where the faces are presented one at a time) is under the control of the police, and thus is a system estimator at the investigative stage. But when that case moves to the court of law, whether a simultaneous or sequential lineup was presented is a variable that is not under the control of the judge and jurors. The judge and jurors simply need to decide how much weight to place on the evidence that has been collected.

2.29.1.5 A New Conceptualization of Variables That Affect Memory

Fig. 2 depicts a different way to conceptualize the variables that affect eyewitness memory, and how, over the course of a crime and its subsequent investigation, the nature of the variables change. Variables are categorized according to its occurrence in a stage in memory and the stage in the crime and subsequent criminal proceedings. The events of the crime are encoded, stored (maintained in memory) during the time between the crime and the reporting of the crime, and finally the memory is retrieved when attempting to identify the perpetrator.

Consider the following example in which a robbery is witnessed. If the perpetrator is of a different race than the witness, then that may affect later identification because of the cross-race bias (**Meissner and Brigham, 2001**), a phenomenon thought to occur during encoding). The time between witnessing the robbery and reporting the crime (storage) would also affect memory because longer retention intervals will decrease memory performance because of forgetting and interference and create opportunities for influences such as misinformation (e.g., **Loftus, 2005**) and suggestibility (e.g., **Zaragoza and Lane, 1994**; **Lindsay and Johnson, 1989**). When the witness attempts to identify the perpetrator from a lineup (retrieval), the type of lineup, such as a

simultaneous lineup, may make a difference in memory performance compared to if memory was tested on, say, a sequential lineup.

2.29.1.6 Two Types of Accuracy

The organizational structure in [Fig. 2](#) provides a road map for the types of analyses that are important to the stakeholders that deal with eyewitness evidence. Both sets of stakeholders (i.e., policymakers and judges/jurors) are interested in accuracy, but in different types of accuracy, and therefore, results from different types of analyses are relevant (Mickes, 2015). One type of accuracy is discriminability. Discriminability is the ability to distinguish the target, or the guilty suspect (the individual who committed the crime), from the innocent suspect. A different type of accuracy—reliability—indicates the likelihood that the identified suspect is guilty for a given level of confidence reported by the eyewitness. Policymakers should consider discriminability when making decisions regarding whether one identification procedure is superior to another. However, courtroom decisions made by judges and jurors, regarding the likely accuracy of witnesses, should consider reliability (Mickes, 2015).

We begin our chapter by expanding upon the assessment of discriminability and reliability. Next, we will turn to a review of several variables known to affect eyewitness memory and will consider the impact of these variables on discriminability and reliability. These variables will be organized according to the memory phase upon which they have the biggest impact (encoding, storage, or retrieval). We will close with the implications: although particular variables impair witness discriminability, witnesses are still often reliable, because identifications made with high confidence on the initial memory test are typically highly accurate.

2.29.2 Measuring Eyewitness Memory

2.29.2.2 Discriminability

2.29.2.2.1 Past Measure of Discriminability: The Diagnosticity Ratio

For many years, claims about discriminability were almost always made on the basis of the diagnosticity ratio (e.g., Steblay et al., 2001, 2011). However, the diagnosticity ratio is not the best way to measure discriminability because it conflates response bias, which is the likelihood of picking someone from a lineup, with discriminability (Wixted and Mickes, 2012; Gronlund et al., 2014). This can lead to misleading conclusions about which procedure maximizes eyewitness performance. ROC analysis and d' are more suitable measures of discriminability (National Research Council, 2014).

2.29.2.2.2 Nonparametric Measure of Discriminability: Receiver Operating Characteristic Analysis

ROC analysis is widely used in other fields, such as diagnostic medicine (Lusted, 1971a,b; Swets, 1988; Swets et al., 2000), and was adapted for use in lineup data (Wixted and Mickes, 2012; Mickes et al., 2012). It separately measures discriminability and response bias. To maximize discriminability, the best procedure is the one that reduces *both* errors that can occur: (1) incorrectly identifying an innocent suspect (false alarm) and (2) failing to identify guilty suspect (miss). An ROC plot is a plot of the correct ID rate and false ID rate for every level of response bias (confidence) and is shown using hypothetical data in Fig. 3. The corresponding data used to plot the ROC curves are shown in the table below Fig. 3 in red font. The bold values represent the overall correct ID rates and false ID rates for both procedures (and the values used to compute diagnosticity ratios). In Fig. 3, the leftmost point is the correct ID rate and false ID

rate reflecting those identifications made with the highest level of confidence (100% confident, the most conservative response bias). The next point to the right is the correct and false ID rate pair for those identifications made with medium or high confidence. The rightmost point is the correct rate and false ID rate for those identifications made with low, medium, or high levels of confidence. In other words, the rightmost point is the overall correct ID rate and false ID rate. The dashed line is the line of chance performance, and if the points fell on that line, it would mean that the eyewitness has no ability to distinguish innocent from guilty suspects. The further the curve falls above the line of chance performance, the better the discriminability. That is, the better eyewitnesses can distinguish innocent from guilty suspects. In [Fig. 3](#), procedure A gives rise to greater discriminability than procedure B, because the ROC is further from the chance line. To measure the statistical differences between the curves, the *partial area under the curve* (pAUC) is computed and compared (as described in detail in [Gronlund et al., 2014](#)). ROC can be considered a measure of “empirical discriminability”; it informs us about the accuracy of a procedure and requires no assumptions about the form of the underlying memory distributions or characteristics of the decision process (Wixted & Mickes, 2018).

ROC analysis for lineup data ([National Research Council \(2014\)](#)) has not been without its critics (e.g., [Lampinen, 2016](#); [Wells et al., 2015a](#)). [Wixted and Mickes \(2015a,b\)](#) responded to these concerns in detail. Recently, Smith et al. (2020) also proposed a different approach—full ROC curves—to measure a new construct called “investigator discriminability”. This analysis does not measure identification decisions by eyewitnesses (which is our focus here) but instead applies to arrest decisions made by police investigators. The new approach can also lead to misleading conclusions about which procedure should be recommended as the ordering of the points is arbitrary, untethered to any formal model of decision-making (Wilson & Colloff, 2020).

When the goal is to measure eyewitness discriminability to inform policymakers, ROC analysis as introduced by Mickes et al. (2012), and endorsed by The US National Academy of Sciences committee, should be used.

2.29.2.2.3 Parametric Measure of Discriminability: d'

d' can also be computed using correct and false ID rates and generally provides a closer approximation to the truth than the diagnosticity ratio in that d' is little affected by response biases (Mickes et al., 2014). d' makes some theoretical assumptions about the underlying data.

Consider a simple signal detection interpretation of a lineup task (top panel of Fig. 4) that maps onto the lineup situation (for a similar instantiation, see Clark, 2003). The model assumes that a witness first determines the lineup member with the greatest memory strength and then identifies that lineup member if his or her memory strength exceeds a response criterion, c (otherwise, the lineup is rejected; known as the Independent Observations model, Wixted et al., 2018). This model assumes Gaussian distributions of memory strength for fillers, innocent suspects, and guilty suspects with means of μ_{Filler} , μ_{Innocent} , and μ_{Guilty} , respectively, and standard deviations of 1 (assuming equal variance). A six-member target-present lineup is conceptualized as five random draws from the filler distribution and one random draw from the guilty distribution. The conceptualization of a target-absent lineup depends on whether the lineup is fair (i.e., the suspect does not stand out among the fillers) or unfair. If the lineup is fair, the distribution for fillers and the innocent suspect are the same, and a six-member target-absent lineup is conceptualized as six random draws from the filler distribution (top panel). However, if

the lineup is unfair, a six-member target-absent lineup is conceptualized as five random draws from the filler distribution and one random draw from the innocent suspect distribution (bottom panel). In all cases, the model assumes that the eyewitness always selects the lineup member with the greatest associated memory strength provided that this strength exceeds a response criterion, c . The ability of eyewitnesses to discriminate innocent from guilty suspects is of interest, which is represented by the distance between the means of the μ_{Innocent} and μ_{Guilty} distributions (i.e., the greater the distance, the greater the discriminability). d' is a measure of the distance between μ_{Innocent} and μ_{Guilty} distributions. It is therefore considered a measure of “psychological discriminability”, which is discriminability in the mind of a participant.

Consider again [Fig. 3](#), which shows the suspect ID rates, filler ID rates, and no IDs for all levels of confidence for target-present and target-absent lineups for the two hypothetical procedures. All three summary statistics (the diagnosticity ratio, pAUC, and d') are presented. In each case, procedure A outperforms procedure B. However, the problem with using the diagnosticity ratio to measure discriminability is clear. For each level of confidence, the diagnosticity ratio can be computed, so for procedure A, the diagnosticity ratio for identifications made with low, medium, and high levels of confidence are 6.8 (.536/.079), 10.0 (.357/.036), and 22.0 (.157/.007), respectively. For procedure B, the diagnosticity ratio for identifications made with low, medium, and high levels of confidence are 2.5 (.257/.103), 3.1 (.143/.046), and 4.0 (.057/.014), respectively. Thus, the diagnosticity ratio increases monotonically as responding becomes more conservative. It is an error to assume that the higher diagnosticity ratio is more accurate (i.e., results in better discriminability) because it is another way of saying that

conservative responding is more accurate. In this example, regardless of response bias—liberal to conservative—discriminability is the same across the ROC curve. The d' values, on the other hand, do not markedly change as responding becomes more conservative. For procedure A, $d' = 1.4$ for identifications made with low confidence and $d' = 1.5$ for identifications made with medium and high levels of confidence. For procedure B, $d' = 0.6$ for identifications made with low, medium, and high levels of confidence. The stability of d' across changing response biases is why d' is a more appropriate measure of discriminability than the diagnosticity ratio.

Therefore, to assess the current state of knowledge (sections [Variables That Affect Encoding](#), [Variables That Affect Storage](#), and [Variables That Affect Retrieval](#)), we compute d' values to assess discriminability in the cases where ROC analysis has yet to be conducted.

2.29.2.3 Reliability

The reliability of an identification is assessed by considering positive predictive value. Positive predictive value is the probability that an identified individual is the perpetrator. The most effective way to measure positive predictive value is to take confidence at the time of the **initial identification** into account. Highly confident witnesses testifying in court are persuasive to jurors ([Penrod and Cutler, 1995](#)), however, this confidence is not necessarily diagnostic of accuracy. There are many powerful forces that can adversely influence the initial identification confidence, such as providing feedback immediately after an identification (e.g., [Wells and Bradfield, 1998](#)), providing misinformation (e.g., [Loftus, 2005](#)), and making suggestive statements (e.g., [Zaragoza and Lane, 1994](#); [Lindsay and Johnson, 1989](#); [Loftus and Palmer, 1974](#)). Thus, confidence expressed at trial is analogous to considering DNA evidence from a

contaminated crime scene (Wixted & Mickes, 2022; Wixted et al., 2018). However, confidence expressed at an initial identification attempt (assuming that the lineup and its administration are fair) is diagnostic of accuracy (Wixted & Wells, 2017). More recently, social media platforms (e.g., Facebook, Instagram) have provided new ways for eyewitnesses to try to find the perpetrator, which may contaminate a subsequent identification procedure administered by the police (which is assumed to be the initial identification attempt). Consequently, the police should instruct witnesses not to search for faces on social media, just as they should instruct witnesses not to speak to one another (Wells et al., 2020 see Paterson et al., 2011), and inquire if a witness did search for faces prior to administering the lineup test. To reiterate, it is the initial confidence report administered by the police that we refer to throughout this chapter, NOT the confidence reported at trial. Moreover, we focus on the accuracy associated with high-confidence IDs assuming equal base rates (i.e., an equal number of target-present and target-absent lineups). This is because (1) analyses of the confidence-accuracy relationship is usually conducted on data from lab experiments, where there are typically a relative equal number of participants viewing target-present and target-absent lineups, and (2) because the surprisingly high levels of accuracy associated with these high-confidence IDs, and the capability of eyewitnesses to compensate for variables that adversely affect discriminability, is central to our claim that eyewitness evidence can be more trustworthy than is typically thought.

2.29.2.3.1 Past Measure of the Confidence–Accuracy Relationship: The Point Biserial

Correlation Coefficient

Early research examining the relationship between confidence and accuracy yielded weak point-biserial correlation coefficients, prompting statements such as, "... the eyewitness confidence-accuracy relation is weak under good laboratory conditions and functionally useless in forensically representative settings" (Wells and Murray, 1984, p. 165; but see Sporer et al., 1995). Yet, Juslin et al. (1996) showed that the point-biserial correlation coefficient was misleading because it can mask the strength of the confidence-accuracy relationship and argued that calibration analysis should be conducted instead. Since then, the work of Brewer and colleagues has shown that the relationship is typically strong, when using the more appropriate calibration analysis (e.g., Brewer and Wells, 2006; Palmer et al., 2013; Sauer et al., 2010), and when the focus is on only choosers (i.e., those who made an identification) at an initial identification attempt. The typically strong relationship between confidence and accuracy became even more clear in subsequent work using confidence-accuracy characteristic analysis (e.g., Mickes, 2015; Wixted and Wells, 2017; Wixted et al., 2015a), leading these researchers to advocate that collecting confidence is the most important factor to determine eyewitness accuracy, and more important than other estimator variables that might exist in a case (e.g., Wixted and Mickes, 2022). We discuss those two ways to measure positive predictive value using confidence, next.

2.29.2.3.2 Calibration Analysis

Calibration analysis measures the relationship between the subjective probability that an identification is correct and the objective probability that the identification is correct. There are

different approaches to computing calibration, and one approach involves including identifications of fillers. This dependent variable is then plotted as accuracy as a function of confidence. The top panel of Fig. 5 shows the calibration plot of the hypothetical data depicted in Fig. 3.

2.29.2.3.3 *Confidence–Accuracy Characteristic Analysis*

A close relative of calibration analysis that only considers suspect identifications is confidence–accuracy characteristic (CAC) analysis (Mickes, 2015). One advantage CAC analysis has over calibration analysis is that the scale does not need to be from 0% to 100%, and instead confidence can be assessed using any type of ordinal scale. More importantly, CAC analysis specifically addresses the question that judges and jurors have (unlike calibration): What is the likelihood that this suspect is guilty (because only suspect IDs, not filler IDs, are advanced to trial)? The bottom panel of Fig. 5 shows the CAC plot of the same hypothetical data as in the top panel. Note that chance accuracy in a CAC plot is 50% if an equal number of target-present and target-absent lineups are used.

We now have laid out how to measure discriminability (ROC analysis, or d' if ROC analyses have not yet been conducted) and how to assess reliability by measuring the relationship between confidence and accuracy (CAC analyses). It is time to turn our attention to prior research involving the many variables that affect eyewitness memory. We organize this discussion by considering whether the primary impact of these variables takes place during the encoding, storage, or retrieval phases of memory.

2.29.3 Variables That Affect Encoding

2.29.3.1 Weapon Focus

Weapon focus refers to a reduction in discriminability for the perpetrator and the details surrounding a crime that involved a weapon (Loftus, 1979; Loftus et al., 1987). The effect likely occurs during encoding because attention is directed to the weapon rather than on the other aspects of the crime or the perpetrator. Two prominent theories, both proposed in the original empirical weapon focus studies, attempt to account for the effect and continue to dominate the literature. These accounts posit that weapon focus is due either to high arousal or to the presence of an unusual object (Loftus et al.). Loftus et al., crediting Easterbrook (1959), proposed that the presence of a weapon causes high arousal that might consequently narrow an eyewitness' attentional focus. They also proposed that lower accuracy might be due to a tendency to fixate on an unusual object. The two possibilities are commonly referred to as the arousal/threat hypothesis and the unusual item hypothesis, respectively. Since then, investigations have tested the predictions of both hypotheses, with some support for each account (e.g., see Davies, Smith, & Blincoe, 2008; Hope & Wright, 2007).

Some of the studies test recall of events and/or test recognition of the perpetrator from a lineup. In many studies, memory was tested only on target-present (e.g., Shaw and Skolnick, 2001; Pickel, 1998) or target-absent lineups (e.g., Hulse and Memon, 2006; Maass and Kohnken, 1989), which is problematic for measuring discriminability (e.g., Mickes and Wixted, 2015; see Rotello et al., 2015).

The first weapon focus experiment using ROC analysis was published in 2014 (Carlson and Carlson, 2014). In this experiment, participants watched a video of a crime in which the perpetrator either did or did not have a weapon or a distinctive feature and then were tested on a sequential or simultaneous lineup. Discriminability was lower in the weapon-present than the

weapon-absent condition when the distinctive feature was not present. Thus, using ROC analysis, the weapon focus effect was found.

What is the effect of weapon focus on the reliability of an ID (the relationship between confidence and accuracy as assessed by CAC analysis)? Regardless of whether a weapon causes lower discriminability, what matters for applied purposes is whether eyewitnesses can take a factor (like weapon presence) into account when reporting their level of confidence (i.e., the CAC results). These are the results that can inform judges and jurors when deciding about defendants' culpability. If accuracy for identifications made with high confidence is the same regardless of the presence or absence of a weapon at encoding, then that is a more relevant consideration than if memory is overall worse if a weapon is present during the crime. We limited the CAC analysis to the weapon and no weapon conditions from [Carlson and Carlson \(2014\)](#) (and excluded the condition in which an artificial feature was added to the target). Our CAC analysis revealed that high-confidence accuracy was very high whether a weapon was present or not (e.g., average accuracy, collapsed across condition, was 97%). In other words, the weapon focus effect does not affect reliability.

In a subsequent investigation of the reliability of the weapon focus effect, participants were assigned to one of three conditions: weapon present, weapon present but concealed, and weapon absent ([Carlson et al., 2016](#)). There was a large discriminability difference between the weapon present and weapon absent and concealed conditions, but again the story was different for reliability. Again, CAC analysis revealed participants who provided the highest levels of confidence (90–100%) were highly accurate (97–99%), regardless of the weapon condition. If this result continues to replicate, these are the results that judges and jurors need to know: at the

high end of the confidence scale, the likelihood that the suspect is guilty (the reliability) is similar, and high, irrespective of the presence of a weapon.

2.29.3.2 Exposure Duration and Divided Attention

The same distinction between discriminability and reliability should be made when it comes to other variables that affect eyewitness identification, such as exposure duration, divided attention, etc. No one would dispute that longer exposure times to stimuli, including a perpetrator, should generally result in greater discriminability than shorter exposure times. Nor would anyone dispute that greater attention paid during encoding should generally lead to greater discriminability. [Palmer et al. \(2013\)](#) conducted experiments in which they manipulated two variables that affect encoding: exposure duration (5 s vs. 90 s, experiment 1); attention (full vs. divided, experiment 2); and one variable that affects retrieval: retention interval (immediately tested vs. tested after a delay, experiment 1). In experiment 1, one experimenter approached potential participants in public places and asked if they would take part in an experiment. If they agreed, a second experimenter would step into view for either 5 s or 90 s; the participants were tested immediately or 6–8 days later. Collapsing across the retention interval conditions, ROC analysis showed that the 90-second encoding condition yielded greater discriminability than the 5-second condition. What about the confidence–accuracy relationship? Palmer et al. provided the data to conduct CAC analysis so that we could answer the question: Did those in the 5-second condition appreciate the fact that they only saw the target briefly and adjust their confidence accordingly? Yes. Identifications made with high confidence had the same accuracy regardless of exposure duration (the CAC results of their experiment are reported in [Mickes, 2015](#)).

In experiment 2, participants watched a video of two target individuals. Participants in the divided attention condition had to respond to low and high tones while watching the video. Those in the control condition were told to ignore the tones while watching the video. Discriminability was lower in the divided attention condition compared to the full attention condition. Nevertheless, identifications made with high confidence were highly accurate for both conditions (over 96%). Once again, participants could adjust their confidence to reflect the likelihood they were making an error with these variables that affect memory at encoding.

2.29.3.3 Cross-Race Bias

The cross-race bias is a phenomenon in which people are more accurate at recognizing faces of individuals of the same race than individuals of another race (Chance and Goldstein, 1981; Malpass and Kravitz, 1969; Meissner and Brigham, 2001). The lower discriminability (i.e., ability to discriminate old from new faces) for other race faces is generally indicated by an increased false alarm rate and occasionally by a reduced hit rate (Meissner and Brigham, 2001). Participants also generally exhibit a more liberal response bias for cross-race faces, meaning that participants are more likely to endorse cross-race faces as previously seen, irrespective of whether they were. The cross-race bias has been found across a variety of conditions and races (e.g., Colloff et al., 2022; Ng and Lindsay, 1994; Platz and Hosch, 1988; Wright et al., 2001).

The cross-race bias is thought to arise during cognitive and social processes at encoding. Bornstein et al. (2013) were unable to reduce the deficit by delivering cautionary instructions at retrieval. Golby et al. (2001) used functional magnetic resonance imaging and found greater fusiform face area activation, the first stage of face-specific processing, for same race faces.

Attempts at explaining the cross-race effect have taken two general approaches. One approach attributes the deficit to differential perceptual experience (e.g., differences in the amount or quality of contact, Brigham and Malpass, 1985; Malpass and Kravitz, 1969), which has resulted in a wide range of suggestions for how differential perceptual experience could be manifested (e.g., Freeman et al., 2016). For example, DeGutis et al. (2013) found that own-race faces were processed more holistically (and therefore more accurately) than cross-race faces. Face space explanations (e.g., Valentine, 1991) documented a more confusable psychological representation of cross-race faces, which would arise if participants focus on the wrong features for discriminating cross-race faces (Papesh and Goldinger, 2010).

The second approach attributes the cross-race deficit to differential social categorization (e.g., Sporer, 2001). Bernstein et al. (2007) demonstrated the power of social categorization to differentially influence memory for in-group (same-race) versus out-group (cross-race) faces. White participants viewed a series of photos of White faces on red and green backgrounds, they were told that the photos on a red background went to their university (in-group) and those on a green background went to their rival university (out-group). Participants showed greater d' for the in-group photos ($d' = 1.23$ vs. 0.94). A control group exhibited no memory difference for the photos ($d' = 1.08$ and 1.14 , for red and green backgrounds). Consequently, it appears that the cross-race effect can be considered a cross-category or an out-group effect.

The most comprehensive current theory, the categorization–individuation model (Hugenberg et al., 2010), ties these two approaches together by proposing that the ability to remember faces that belong to a different category/race is a function of three factors: social categorization, motivated individuation, and perceptual experience. According to this theory, the out-group bias

arises from the tendency to attend to *identity*-diagnostic (individualistic) features of in-group individuals but *category*-diagnostic features (e.g., skin tone) of out-group individuals. This theory also provides explanations regarding factors that create, diminish, or eliminate an out-group memory deficit (e.g., see Ackerman et al., 2006; Shriver and Hugenberg, 2010).

Hourihan et al. (2012) found that participants are worse at judging whether they will subsequently remember a cross-race face. This has potential implications for the reliance on eyewitness confidence when a cross-race ID is involved. If eyewitnesses are not aware of this difficulty, they may not be able to adjust their post-identification confidence judgment accordingly. But the eyewitness literature finds that confidence generally appears to be informative about identification accuracy, for both own-race and other-race identifications. Dodson and Dobolyi (2015) showed that confidence was very informative of accuracy and race was minimally important (the accuracy for high-confidence same race and cross-race IDs both exceeded 95%). Again, these participants were able to adjust the likelihood that they made a correct ID even though they were more likely to make an error when making a cross-race ID. A similar finding was reported by Nguyen et al. (2017) (for experiments in which performance was above chance levels) who reanalyzed data from four cross-race face recognition experiments, and by Colloff et al. (2022, experiment 1), who tested participants on both photo lineups and novel interactive lineups where the lineup faces could be moved to be viewed from different angles.

One consideration is that in these studies participants were tested on their memory for multiple faces, with multiple lineups, and therefore could have learnt over multiple trials how to adjust their confidence judgments appropriately. In real life, witnesses typically view one crime and one lineup per perpetrator. In a single-trial experiment by Colloff et al. (2022, experiment 2),

participants saw one perpetrator in a mock-crime video and saw one lineup. Here, high-confidence cross-race identifications were less accurate than high-confidence own-race identifications. Another factor that differed across experiment 1 and 2 of Colloff et al. (2022) was previous contact with cross-race faces, which was likely higher in experiment 1 (where high-confidence own and cross-race identifications were equally reliable) compared to experiment 2 (where they were not). White and South Asian participants in experiment 1 were students recruited from a UK university in a multicultural city, whereas experiment 2 recruited participants from the general population in the UK or US (White sample) or South Asian countries (predominantly India) where contact with cross-race faces was likely to be less prevalent, on average. More research is required to determine the effect of trial type, and cross-race contact/expertise on cross-race reliability. Nevertheless, for both own- and cross-race identifications across studies, increased confidence is associated with increased suspect identification accuracy, highlighting that confidence can be a useful (but perhaps an imperfect) indicator of likely accuracy.

2.29.3.4 Stress and Arousal

Arousal signifies general physiological and psychological activation; stress arises from an imbalance between the physical and psychological demands and the ability to respond (Hoscheidt et al., 2014). Stress and arousal have complex effects on memory. For example, stress appears to impair memory for neutral information (Payne et al., 2002) but enhance memory for emotionally arousing materials (Buchanan and Lovallo, 2001). The research conducted on lineups focuses on discriminability; more research needs to be completed assessing reliability using CAC analysis. Crime can never be made less stressful, but we need to determine if eyewitnesses are able to compensate for that stress when they assess their confidence.

A meta-analysis by Deffenbacher et al. (2004) concluded that high levels of stress adversely impacts eyewitness IDs of a target. Across 27 tests, correct ID accuracy was 0.42 versus 0.54 for high stress and low stress, respectively. Interestingly, increased stress did not adversely affect target-absent accuracy. Valentine and Mesout (2008) tested individuals visiting the London Dungeon who were exposed to an individual during the tour that they later were asked to ID from a target-present lineup. Valentine and Mesout classified participants based on a median split based on their anxiety score in reaction to the Dungeon performance. Among the low anxious individuals, 75% chose the target, but only 18% of the high anxious individuals chose the target. Unfortunately, without a target-absent lineup for comparison, and given large differences in response biases across the two groups (high anxious more conservative), conclusions are tentative.

Morgan et al. (2004) conducted the most compelling experiment examining memory for a perpetrator. It is the most compelling because it involves far greater levels of stress than can ethically be used in the laboratory. Participants were soldiers undergoing prisoner-of-war training, which involved sleep and food deprivation. After a week of classroom training, participants were confined in a mock prisoner of war camp, and while in isolation were subjected to low-stress and high-stress interrogations. Participants were threatened with physical violence in the high-stress interrogation. ID attempts were made approximately 24 h after completion of the training. Despite interacting with their interrogators for more than 30 min, participants had great difficulty subsequently identifying an interrogator (some viewed a live lineup, others a photo lineup). We took the average of the live and photo lineup data and found that d' was greater (and the criterion more liberal) when identifying the low-stress interrogator ($d' = 2.33$, $c = 0.66$) than the high-stress interrogator ($d' = 1.36$, $c = 1.20$). Morgan et al. (2013) used a similar

sample of participants undergoing prisoner-of-war training. All participants experienced high-stress interrogations but varied in whether, and how, they were exposed to misinformation. Some groups subsequently viewed a nine-person target-absent lineup, from which 53% of the control participants made a false positive selection. Those individuals in the photographic misinformation condition (who were exposed to a filler photo and asked a series of questions about the interrogation while viewing this photo) had a false ID rate of 91% and, unsurprisingly, most of these selections were of the filler.

In another investigation of the effect of stress on lineup performance, participants watched a video of a mock crime while either being exposed to a stressor (cold pressor) or not, and later memory for the target in the video was tested from a lineup (Sauerland et al., 2016). Cortisol levels validated the efficacy of the stress manipulation, yet neither ROC analysis nor CAC analysis yielded a difference between the groups. Identifications made with high confidence were high in accuracy (i.e., over 95% correct) regardless of level of stress. The authors acknowledged that the sample size was small, so more research needs to be conducted in this area. Marr et al. (2021) also found no evidence that acute stress during encoding or retrieval affected face (or word) recognition performance.

In the Houston Police Department field study, over 300 real eyewitnesses to robberies who were faced with a lineup decision made confidence judgments in their identifications (Wells, 2014). Many of these witnesses were victims and thus likely had experienced some degree of stress during encoding of the crime. Despite that, the authors found that the identifications made with high confidence were associated with high accuracy and identifications made with low accuracy were associated with low accuracy (Wixted et al., 2016), suggesting that these eyewitnesses were reliable. The analysis of this field data required models fit to the data to

estimate which lineups were ‘target-present’ and ‘target-absent’ (i.e., which suspect IDs were of guilty versus innocent suspects). Although two different models fit the data well and yielded the same conclusion, the ground truth of the accuracy of the suspect IDs can never be certain.

In sum, stress appears to adversely impact discriminability for the target of an event, be that a mock perpetrator, an actor in the London Dungeon, or an interrogator. If the police could control the level of stress experienced by eyewitnesses at the time of a crime, then they should ensure low stress to ensure higher discriminability during a later identification. But, of course, they cannot, so the key question for the legal system concerns the effect of stress on reliability. Initial CAC results and implications of the Houston Police Department field study suggest that stress does not impair reliability (i.e., stress does not cause eyewitnesses to mistakenly identify innocent suspects with high confidence). If confirmed by additional research, this would be a fact that judges and jurors should be made aware of.

2.29.4 Variables That Affect Storage

2.29.4.1 Verbal Overshadowing

The verbal overshadowing effect posits that memory is adversely affected after providing a verbal description of a previously presented stimulus (e.g., a face). **Schooler and Engstler-Schooler (1990)** coined the term and found in a series of experiments that those who verbally described a perpetrator from a mock-crime video were less able to correctly identify the perpetrator from a later lineup than those in the control condition who did not verbally describe the perpetrator. In experiment 1 of the original paper, participants were tested on target-present lineup procedures, and the correct ID rate was 0.64 in the control condition compared to only 0.37 in the verbal description condition. This counterintuitive finding sparked much interest and

follow-up research: the original paper has been cited 1308 times (Google Scholar search retrieved October 30, 2023). Note that no conclusion can be reached regarding discriminability without a target-absent condition.

The follow-up research does not tell a consistent story. The lack of consistency was the impetus for a meta-analysis conducted by Meissner and Brigham (2001). Based on results from 29 investigations of the verbal overshadowing effect the authors concluded that the effect was real but small. Other investigations were conducted on list learning experiments in which the verbalization conditions fared better than the control conditions. The change in paradigms (i.e., list learning vs. forensically relevant experiments) might account for the inconsistencies reported since the original verbal overshadowing effect paper (which used a forensically relevant one-trial paradigm).

There are three main hypotheses to explain the verbal overshadowing effect. The content account (e.g., Meissner and Brigham, 2001) holds that the verbal description interferes with the memory of the target, causing a reduction in discriminability (i.e., the ability to distinguish between the innocent and guilty suspect). The criterion-shift account (Clare and Lewandowsky, 2004) holds that verbal overshadowing reflects a change in response bias (i.e., the likelihood to choose someone from a lineup) rather than a change in discriminability. The processing account holds that the switch from visual to verbal processing (Schooler, 2002) affects both discriminability *and* response bias (Chin and Schooler, 2008).

Because of the lack of consistent findings, and because the importance of conducting direct replications has been highlighted in the field of psychology (Pashler and Wagenmakers, 2012), experiments 1 and 4 in the original Schooler and Engstler-Schooler (1990) paper were chosen as one of the first preregistered replication report projects (Open Science Collaboration, 2015).

Independent laboratories responded to the call to attempt to replicate the finding, and the meta-analysis from that work can inform us about the true size of the verbal overshadowing effect (Alogna et al., 2014). The results revealed a small, but significant, verbal overshadowing deficit, which was larger if the description task was separated in time from the original event.

In the original paper, participants were only tested on target-present lineups. But to get a complete picture of how verbalization affects participants' ability to discriminate innocent from guilty suspects, one needs to also measure false ID rates (in target-absent lineups). Despite replicating the original result (Alogna et al., 2014), because there was no way to measure false IDs, the effect verbalization has on identification performance (discriminability) remained unclear (Mickes, 2016; Mickes and Wixted, 2015; Rotello et al., 2015).

Since the replication study by Alona et al. (2014), four experiments investigating the effects of verbal overshadowing were conducted with both target-present and -absent conditions that allowed ROC and CAC analyses (Wilson et al., 2018). For both lineups and showups, when the description occurred immediately after the study phase and before a distractor task, the ROC curves were not significantly different (i.e., discriminability was not different), but when the description occurred 20 minutes after the study phase (i.e., after the distractor task) and immediately before the test phase, the ROC curves revealed that discriminability was lower in the verbalization condition. That is, a verbal overshadowing effect was observed. Wilson et al., (2018) suggested that the differences due to the timing of the descriptions exist because immediate verbalizations contain more diagnostic descriptions. They found that participants provided less diagnostic descriptions after a delay (presumably due to forgetting of more specific diagnostic details). This was verified by other participants who did not view the mock crime

video but were provided with those descriptions. Those participants given descriptions that had been written immediately after encoding (i.e., the more diagnostic descriptions) were better able to identify the perpetrator than those participants given the descriptions that had been written after a delay (see also Marmurek et al., 2021). Importantly, in all experiments conducted by Wilson et al. (i.e., even when a verbal overshadowing effect was observed on discriminability), CAC analysis revealed that high-confidence identifications were similarly reliable, and highly accurate, across the verbalization and control conditions (the pattern of results was replicated in Holdstock et al., 2022). The CAC results are reassuring given that the criminal justice system relies on verbal reports of crimes.

2.29.4.2 Retention Interval

What do judges and jurors need to know about increasing retention interval that they do not already know? It is no surprise that memory gets worse with time (e.g., Juslin et al., 1996; Palmer et al., 2013; Read et al., 1998; Sauer et al., 2010). But what about reliability? Wixted et al. (2016) reanalyzed data from four retention interval studies where participants were tested either immediately or 1 week after the encoding event (Juslin et al., 1996; Palmer et al., 2013), immediately or 3 weeks after (Sauer et al., 2010), and 3 months or 6-to-9 months later (Read et al., 1998). Remarkably, the CAC results showed that regardless of retention interval, identifications made with high confidence were highly accurate, even after 9 months. In each case, discriminability (as measured by d') decreased as retention interval increased (see also Morgan et al., 2019, for description of a sleep experiment in which there were differences in discriminability after a delay of one day but the delay did not cause a difference in CACs). Thus, despite ever-decreasing discriminability with time, high-confidence identifications remain highly

accurate, suggesting that a confident witness could be relied upon to make an accurate suspect identification even after a delay.

2.29.4.3 Suggestibility Effects

It is well known that memory is reconstructive (e.g., Bartlett, 1932; Roediger and McDermott, 1995), and that post event suggestions are often incorporated and reported as part of an original event (e.g., Loftus and Palmer, 1974). This occurs, in part, due to poor source monitoring (Johnson et al., 1993). Eyewitnesses become confused between their own experiences and inferences they draw (Do I remember the tattoo or did I infer that the gang member who robbed me had a tattoo?), between what they experience and what a co-witness might report (Did you see the scar on the robber's face?), or what they might see or hear in the media.

Bonham and González-Vallejo (2009) examined the effect of misinformation on eyewitness discriminability and calibration. Participants watched a mock crime video and then responded to questionnaires or read narratives that contained correct information and misinformation about what transpired in the video. Discriminability was poorer for details regarding the crime when misinformation was introduced. Likewise, the confidence–accuracy relationship suffered as a function of misinformation. In other words, individuals incorrectly maintained high confidence in the accuracy of memories about which they had been misinformed. The same pattern has been observed in witness interviews. In a recent study, participants became highly overconfident in the accuracy of their memories when they had been unknowingly exposed to misinformation (Spearing & Wade, 2022). Unlike factors like divided attention and retention interval, eyewitnesses appear to be unable to adjust their confidence to reflect the reduced accuracy that arises from misinformation. As a result, some researchers urge

caution about using confidence to assess eyewitness accuracy (e.g., Berkowitz et al., 2020; Sauer et al., 2019; Wade et al., 2018), arguing that it is difficult to ensure that witnesses' memories in the real world are not contaminated (e.g., with misinformation) before they make an identification decision. However, as described before, the Houston Police Department field study recorded over 300 eyewitness photo lineup decisions (Wells, 2014) and most suspect identifications were made with high confidence, whereas most filler identifications were made with low confidence (Wixted et al., 2015b). High-confidence suspect identifications were estimated to be 96% correct and low-confidence suspect identifications were around 50% correct. This field study provides evidence that highly confident suspect identifications can be accurate in the real world, despite the variables that may influence memory encoding and storage. Though again, note that the ground truth can never be certain in field studies (i.e., we can never be certain that the suspect was guilty), and more research in other jurisdictions is also needed.

2.29.5 Variables That Affect Retrieval

2.29.5.1 Creating Fair Lineups

What constitutes a fair lineup and how a fair lineup is created has much to do with the fillers selected to be in that lineup. In fact, Wogalter et al. (2004) argued that part of the reason why there are fillers is to protect the innocent suspect from being wrongfully identified.

2.29.5.1.1 Filler Selection

What should the characteristics of the fillers be? Should the fillers in the lineup match the witness's verbal description of the perpetrator or should they match how the suspect (who may

be innocent or guilty) looks? And how closely should the fillers match? How should perpetrators with distinctive characteristics be matched? These are questions that have been the focus of many investigations, because knowing the answers to these questions can result in the preparation of the fairest lineup and the resulting best quality eyewitness evidence.

2.29.5.1.1.1 Description-Matched or Suspect-Matched?

Wells et al. (1993) tested the hypothesis that the description-matched method of filler selection would result in greater discriminability than the suspect-matched method of filler selection. Indeed, the description-matched condition yielded a higher d' than the suspect-matched condition (1.61 vs. 0.37, respectively). They concluded, "A good lineup appears to be one in which all lineup members match the eyewitness's pre-lineup description of the culprit but otherwise do not resemble each other." (p. 844). Although follow-up research has replicated a benefit for description-matched versus suspect-matched lineups (e.g., Juslin et al., 1996; Tunnicliff and Clark, 2000), other studies found greater discriminability for the suspect-matched conditions (e.g., Lindsay et al., 1994; Tunnicliff and Clark, 2000; Darling et al., 2007). Clark et al. (2013b), or found no difference and recommended a combination of description- and suspect-matched selection, which is what many police departments appear to do (Bennett et al., in prep; Fitzgerald et al., 2021; Wise et al., 2011). A more recent large-scale study conducting ROC and CAC analysis showed that description-matched fillers enhanced discriminability compared to appearance-matched fillers, cautioning that discriminability decreases as the fillers become too similar to each other and the suspect. Importantly, identifications still were similarly reliable across the suspect-matched and appearance-matched conditions (Carlson et al., 2019).

The research has yielded inconsistent results about which method best supports discriminability, possibly because it has not been guided by a strong theoretical framework in which to make predictions and interpret results (Colloff et al., 2021; Roughton et al., 2023). Colloff et al. (2021) introduced a formal feature-matching model to make predictions about the influence of suspect-filler similarity on eyewitness discriminability. The model predicted that a two-step process would best optimize lineup performance. In Step One a pool of description-matched fillers is collated, all of whom would be suitable for creating a fair lineup because they would be reasonably described in the same way the witness described the perpetrator. In Step Two the goal is to *maximize dissimilarity* by choosing from that pool the fillers who are the least similar-looking to the suspect. The model predicted that this strategy would increase correct identifications of perpetrators without affecting the false-alarm rate of innocent suspects, and therefore increase discriminability. Colloff et al. (2021) confirmed the model's predictions in a large-scale experiment testing over 10,000 participants, suggesting the two-step process for selecting fillers might optimize eyewitness performance (see also Shen et al., 2023, who replicated these results). Notably, this is the same pattern of results predicted and observed by Wells et al. in 1993 without the use of a formal model. Such research suggests that eyewitness discriminability might be optimized by choosing fillers that match the suspect on basic facial features described by the eyewitness (e.g., age, race, gender), but who otherwise are maximally *dissimilar* to the suspect. More research is needed, however, before recommendations for practice are made.

2.29.5.1.1.2 Distinctive Features

Distinctive features, such as a face tattoo or scars, may be encoded during the crime, but how those features are subsequently handled is an issue for law enforcement. There are several propositions to deal with distinctive features when assembling a lineup, and they are as follows: conceal the feature on the suspect's face and conceal the same area on the faces of the fillers, replicate the suspect's feature on the fillers, or leave alone (i.e., do nothing to prevent the distinctive suspect from standing out). Currently, guidelines leave it to the discretion of the identification officer whether they replicate or conceal the feature (PACE Code D; [Technical Working Group for Eyewitness Evidence, 1999](#)). In practice, concealment is often achieved by covering the area with a black block or pixelating the area of the feature.

[Colloff et al. \(2016\)](#) collected data from 9,841 participants to compare replication, block concealment, pixelation concealment, or doing nothing. Participants viewed a video of a crime (one of four videos with four different targets) in which there was a prominent distinctive feature (each target had a different distinctive feature). ROC analysis revealed that discriminability was similar for the replication, block, and pixelation conditions, but significantly lower for the do-nothing condition. This result was replicated in participants in young, middle, and older adulthood ([Colloff et al., 2017](#)) and implies that when constructing the lineup, ignoring the suspect's distinctive feature and leaving a distinctive suspect to stand out is bad practice and any other alternative would be preferred.

The other (i.e., fair) alternatives also lead to better reliability ([Colloff et al., 2016, 2017](#)). At all levels of confidence, accuracy was significantly lower in the do-nothing condition. Moreover, at high levels of confidence (i.e., identifications made with 90%–100% confidence), accuracy in the do-nothing condition was only approximately 60%, whereas in the other conditions, accuracy

was approximately 85%. This finding implies that judges and jurors need to know how the lineup was constructed. For example, if the defendant was identified with high confidence from a lineup in which there was a suspect with a distinctive feature that was not accounted for by replicating, blocking, or pixelating, then even an identification made with high confidence is less trustworthy.

More recently, research has compared how exactly a feature could be replicated across the fillers. The exact feature could be replicated across the lineup fillers (high feature-similarity replication, e.g., the suspect's tribal tattoo could be copied exactly onto the other lineup faces), or the feature could be replicated with variation across the lineup fillers (e.g., a slightly different tribal tattoo on each lineup face), but within the constraints of the witness's description (low feature-similarity replication). Mahmood et al. (2022) found that low feature-similarity replications improve discriminability compared to high feature-similarity replications or doing nothing. These findings are consistent with the recent filler similarity research described above (and the feature-matching model; Colloff et al., 2021; Shen et al., 2023), and suggest that the police should not needlessly match features across a lineup that would otherwise be useful cues to identity, as this is likely to harm discriminability.

2.29.5.2 Identification Procedure/Presentation

2.29.5.2.1 Simultaneous Versus Sequential Lineups

How should lineup members be presented to the eyewitness? In a simultaneous lineup, all lineup members are presented at once and the eyewitness makes only one decision (Which person, if anyone, is the perpetrator?). In a sequential lineup, lineup members are viewed one at a time, and a decision may be required regarding each member (Is this the perpetrator?) before

lineup next member is presented, or a decision is required at the end once the witness has seen the full sequence of lineup members. In the first investigation comparing simultaneous versus sequential photo lineups (Lindsay and Wells, 1985), the false ID rate was much higher in simultaneous lineups than sequential lineups (0.43 vs. 0.17, respectively). The correct ID rate was not much higher in simultaneous lineups than sequential lineups (0.58 vs. 0.50, respectively). Over time, with the exception of two experiments (Lindsay et al., 1991a,b), the differences in false ID rates were not as high as in the original study, and the sequential advantage sometimes failed to replicate (e.g., Carlson et al., 2008; Gronlund et al., 2009). Others concluded that the pattern of results was more consistent with the sequential lineup inducing a conservative criterion shift (e.g., Ebbesen and Flowe, 2002; Meissner et al., 2005; Palmer and Brewer, 2012). Despite this contrary evidence, the claim made was that the sequential lineup was superior (Stebay et al., 2001, 2011). The claims were based on higher diagnosticity ratios for the sequential lineups, and because of these claims, sequential lineups were recommended for use over simultaneous lineups (e.g., Lindsay, 1999; Innocence Project, 2009; Wells et al., 2000). As a result, 30% of jurisdictions across the US switched from using the simultaneous lineups to sequential lineups (Police Executive Research Forum, 2013).

After the importation of ROC analysis to use for lineup data in 2012, the simultaneous lineup has been found to outperform sequential lineups (e.g., Carlson and Carlson, 2014; Dobolyi and Dodson, 2013; Gronlund et al., 2012; Mickes et al., 2012; Seale-Carlisle and Mickes, 2016; see Seale-Carlisle et al., 2019 for a mini-meta analysis; but also see Kaesler et al., 2020; Meyer et al., 2023). One theory explaining these results is diagnostic feature-detection theory (Wixted and Mickes, 2014), which holds that when faces are presented simultaneously, it is easier to detect

(and then discount) non-diagnostic facial features (e.g., race) that are shared by multiple lineup members than when faces are presented sequentially. Discounting non-diagnostic facial features improves discriminability as diagnostic features (i.e., those unique to the perpetrator) are given more weight in the witness's decision. Research has also shown that when participants viewing a sequential lineup are required to make an identification decision to each face before seeing the next face, the ROC falls (i.e., empirical discriminability decreases) as the position of the suspect in the lineup increases (Wilson et al., 2019, see also Rotello and Chen, 2016). This also illustrates the deleterious position effects observed for sequential lineups. Research should further examine the influence of different sequential procedures and decision rules (e.g., when a single decision is made at the end of the sequential lineup compared to when a yes/no decision is made to each face) as this may influence performance.

The US National Academy of Sciences committee and the US Department of Justice decided not to endorse one procedure over the other because they deemed the matter still unresolved (National Research Council, 2014; Yates, 2017). Nonetheless, the Innocence Project ceased to support sequential lineups as a reform, although some US state innocence projects continued to support sequential lineups (e.g., Minnesota) and the sequential lineup is still used in other countries around the world (e.g., UK; Police and Criminal Evidence Act, 1984, last updated in 2017). Two field studies found that simultaneous lineups yield better discriminability than sequential lineups (Amendola and Wixted, 2015; Wixted et al., 2015b), which further bolsters support for simultaneous lineups. Thus, if scientists have a deep theoretical understanding of sequential versus simultaneous lineups, including the different variations of sequential lineups

(e.g., Kaesler et al., under review), and the simultaneous advantage continues to replicate, then policymakers should begin recommending the use of the superior lineup procedure.

Regardless of the lineup procedure that is used, judges and jurors need to know about the reliability of an identification. This issue has been addressed by conducting CAC analyses on experiments that reported ROC analyses. It has been shown that simultaneous and sequential lineups yield equal reliability for identifications made with high confidence (Dobolyi and Dodson, 2013; Gronlund et al., 2012; Experiments 1A and 1B of Mickes et al., 2012; Weber and Brewer, 2004). Thus, while discriminability tends to suffer due to sequential lineup presentation, reliability (at the highest confidence level) is not different.

2.29.5.2.2 Lineup Modality

Historically, lineups globally were conducted “live”, wherein the lineup members are physically present before the witness, but photo and video lineups are now commonplace. In a photo lineup, each lineup member is shown in a static image from the shoulders up facing the camera (like a passport photo), and in a video lineup each lineup member moves their head from left to right and then back to face the camera. Photo lineups are the most common procedure used in the US, and video lineups are the recommended procedure in the UK (Police and Criminal Evidence Act, 1984). Research suggests that live lineups do not improve discriminability compared to video lineups (Fitzgerald et al., 2018; Rubínová et al., 2021). Moreover, there appears to be no benefit of using videos compared to photos on discriminability (Meyer et al., 2023; Rubínová et al., 2021; Seale-Carlisle et al., 2019). However, some results are mixed across studies (e.g., see Cutler & Fisher, 1990; Seale-Carlisle & Mickes, 2016; Valentine et al., 2007). Nevertheless, because live lineups are difficult to organize, some

researchers have recommended that photo and video be used as they are more practical, with no clear detrimental effects on witness performance (Fitzgerald et al., 2018).

From a theoretical standpoint, however, it is surprising that discriminability does not vary across procedures. Previous research has not considered the angle from which a witness encoded the perpetrator (Colloff et al., 2021). If a witness saw the side-profile of a perpetrator's face as they entered a building, it is likely that seeing that viewpoint at test would benefit discriminability, as predicted by one of the most influential principles of memory: encoding specificity (Tulving & Thomson, 1973). Profile views are more readily available in video and live lineups, but not in frontal view photo lineups, and therefore more research is needed.

Relatedly, scientists have long been calling for researchers to innovate new procedures, based on psychological theory, that could improve witness discriminability more than current procedures (e.g., [National Research Council, 2014](#), Wells et al., 2006). Recently, researchers have developed and tested a new interactive lineup, in which witnesses can freely rotate the lineup faces to view them from multiple angles. This research finds that discriminability is significantly higher when the viewing angle of the lineup members is the same rather than different from the perpetrator at encoding. Participants also naturally reinstated the angle of the interactive faces to match the angle in which they encoded the perpetrator (Colloff et al., 2021). This highlights the importance of encoding specificity for eyewitness identification (e.g. Tulving & Thomson, 1973), and demonstrates that participants actively seek out cues in the testing environment that match the study environment, and this aids memory retrieval. Research additionally finds that lineups presented in an interactive format improve discriminability compared to photo lineups (Colloff et al., 2022), and simultaneous interactive lineups improve discriminability compared to simultaneous photo and sequential video lineups (where witnesses

make one ID decision at the end of the face sequence), regardless of whether the study faces were encoded from the front or the profile (Meyer et al., 2023). Using modelling, Meyer et al. proposed that interactive lineups increase discriminability because they afford the witness the opportunity to view more diagnostic features, and thereby, non-diagnostic features play a proportionally lesser role. Therefore, witness discriminability may be improved when lineup procedures create retrieval conditions that enable witnesses to actively explore faces and more effectively sample facial features, such as in an interactive procedure.

2.29.5.2.3 Showup Procedures

Unlike a lineup, a showup involves the presentation of a single suspect (not accompanied by fillers) to the eyewitness. Showups have been criticized for their inherently suggestive nature because the eyewitnesses obviously know that the person presented is the police suspect (Goodsell et al., 2013; Steblay et al., 2003). Despite this, and other potential downsides, the showup will likely remain a common way of testing eyewitness memory given that it is an easy way to test memory soon after a crime has occurred. If a suspect who fits the witness's description has been found in the vicinity of a crime, the choice is often between conducting a showup or setting the detained person free.

Simultaneous lineups yield greater discriminability than showups when measured with ROC analysis (Neuschatz et al., 2016; Wetmore et al., 2015; Mickes, 2015; Gronlund et al., 2012). The CAC results differ as well. Showup procedures yield lower reliability even for the identifications made with high confidence compared to simultaneous lineups (Wetmore et al., 2015; Mickes, 2015). Thus, there is a growing body of empirical evidence suggesting that simultaneous lineups

are superior to showups in terms of both discriminability and reliability. If the results continue to replicate, this suggests that lineups should be used when possible.

Although scientists agree that lineups produce better outcomes than do showups (e.g., Wells et al., 2020), there is ongoing debate as to why lineups are better than showups. Because incorrect positive identifications always land on the innocent suspect in a showup, fair fillers in lineups offer some protection against misidentifications of the innocent suspect (called *filler siphoning*; e.g., Wells, 2001). In addition, other researchers have concluded that fair fillers also help witnesses to decide which facial features are useful for the identification decision, in accordance with diagnostic feature-detection theory (Wixted and Mickes, 2014). To test these two accounts, Colloff and Wixted (2020) conducted three experiments and found that presenting similar-looking faces alongside the suspect, even if those similar-looking faces could not be identified, enhanced subjects' ability to discriminate between innocent and guilty suspects compared with presenting the suspect alone (i.e., in a showup). These results suggest that the discriminability advantage in simultaneous lineups is due the useful comparison of multiple faces as predicted by diagnostic-feature-detection theory.

2.29.5.3 Blind Administration

The lineup administrator may—advertently or inadvertently—influence an eyewitness in two ways. (1) The administrator can influence *who* an eyewitness chooses, or *if* an eyewitness chooses. (2) The administrator can provide feedback about a choice that an eyewitness makes. The first source of influence maps onto discriminability (who) and bias (if), respectively. A discriminability influence results from what Greathouse and Kovera (2009) referred to as

steering, whereby the administrator directs a witness toward a particular suspect. Alternatively, bias is affected if an administrator exerts a nonspecific influence that affects the rate at which an eyewitness chooses. There is evidence in the literature for both these patterns (Clark et al., 2009; Greathouse and Kovera, 2009; Haw and Fisher, 2004; Kovera and Evelo, 2017). The second source of influence arises once an eyewitness has made a choice, and the administrator provides post-ID feedback regarding the correctness of that choice (e.g., Wells and Bradfield, 1998; Wright and Skagerberg, 2007; see meta-analysis by Bradfield Douglass and Steblay, 2006). Post-ID feedback tends to inflate eyewitness confidence, as well as inflate several other indices reflecting an eyewitness' memory for the perpetrator (e.g., how long a look they got, how close they were). Researchers have noted the importance of focusing on the information at the time of the initial identification only (i.e., the confidence judgment and ID decision made on the first memory test; e.g., Wixted and Mickes, 2022; Wixted et al., 2015a). If the legal system did this, then post-ID feedback need not be of great concern. But double-blind administration remains important so that the reported confidence associated with the first, fair test of an eyewitness' memory remains unspoiled.

What do we know about the effects of blind administration on lineup discriminability and response bias? There is limited research within the eyewitness domain that compares double- to single-blind lineup administration using target-present and target-absent lineups. Clark et al. (2015) reviewed blind administration and also mention the relative lack of empirical data that focuses on the crucial independent variables (blind vs. nonblind, both target-present and target-absent, without contamination from factors like different types of instructions, lineup presentation methods). We also note that more studies need to be conducted using skilled lineup

administrators (e.g., Clark et al., 2013a). Administrator effects are clearly going to be influenced by the skill (or lack thereof) of the administrators involved (Russano et al., 2006).

One study, conducted by Greathouse and Kovera (2009), did vary lineup presentation, target presence, single- or double-blind, and biased or unbiased lineup instructions. We focus on the simultaneous lineup data, and computed d' and c from the reported correct and false IDs rates. If biased instructions (which implied the suspect was in the lineup and the witness should identify him or her) were employed, discriminability was greater with double-blind than single-blind administration ($d' = 2.19$ vs. $d' = 1.52$, respectively), and double-blind administration induced greater response conservatism ($c = 0.74$ vs. $c = -0.32$, respectively). However, for unbiased instructions (the suspect may or may not be in the lineup), discriminability was slightly greater for single-blind than double-blind administration ($d' = 1.01$ vs. $d' = 0.70$, respectively), and no difference arose in response bias ($c = 0.58$ vs. $c = 0.53$, respectively). These data suggest that combining double-blind control with unbiased instructions might not enhance eyewitness performance.

Although lineup research specifically does not show that performance is enhanced when a blind administrator is used, the National Academy of Sciences report (National Research Council, 2014) and the American Psychology-Law Society (AP-LS) White Paper written by leading experts (Wells et al., 2020) recommend double-blind lineup administration. The reason for this is twofold. First, because other lines of scientific research indicate could happen when testing is not double-blind. A long history of research in psychology and medicine support the merits of double-blind testing (Rosenthal, 2002). Second, due to considerations of procedural fairness. As Clark et al. (2015) argued, although a key goal of eyewitness reforms is to determine

which procedures can enhance discriminability, there also are issues of fairness to consider (e.g., procedural justice, Tyler, 2003). Positive eyewitness identifications cannot be considered as providing independent evidence of guilt if the ID arises from the pressures of a lineup administrator (also see Hasel and Kassin, 2009). Unfortunately, little research has explored the effects of blind lineup administration on reliability, although Clark et al. (2013a) reported that post-ID confidence was lower for those participants who were pushed or steered.

2.29.5.4 Lineup and Juror Instructions

2.29.5.4.1 Lineup Instructions

Numerous attempts have been made to improve eyewitness evidence through instructions, with little success. Some of this research has focused on the instructions given to eyewitnesses, and other research has focused on the instructions given to jurors regarding how to weigh eyewitness evidence. We begin with biased versus unbiased instructions to eyewitnesses.

2.29.5.4.1.1 Biased Versus Unbiased Instructions

Biased instructions imply that the perpetrator is in the lineup, unbiased instructions do not (e.g., “The perpetrator *may or may not* be present”). Malpass and Devine (1981) conducted the first study comparing biased and unbiased instructions and found a d' advantage of 0.75 for unbiased instructions. Research that followed was in general agreement (e.g., Cutler et al., 1987; O'Rourke et al., 1989), which culminated in a meta-analysis in which Steblay (1997) concluded that unbiased instructions decreased choosing from target-absent lineups without decreasing correct IDs from target-present lineups. A large-scale study using ROC analysis found that biased instructions (in the liberal or conservative direction) seemed to impair discriminability compared to unbiased instructions (Mickes et al., 2017). Yet, a meta-analysis by Clark et al. in

2014 showed a different conclusion - that the discriminability advantage for unbiased instructions aggregated across 23 studies was nonexistent ($d' = -.02$). Note that that Mickes et al. (2017) paper was not included in this meta-analysis. In light of these results, it appears that biased versus unbiased instructions largely effect response bias (eyewitnesses are more conservative after receiving unbiased instructions), and might also improve discriminability.

Despite unbiased instructions not consistently being found to enhance discriminability, the National Academy of Sciences report (National Research Council, 2014) nevertheless recommended that they be used: "Witnesses should be instructed that the perpetrator may or may not be in the photo array or lineup..." (p. 73). The AP-LS White paper also recommended that witnesses should be told information to prevent them from being predisposed to identify someone, including that the culprit may not be in the lineup (Wells et al., 2020). To understand the rationale for this recommendation, one must consider the trade-offs between the costs of errors (ID an innocent suspect, the failure to ID a guilty suspect) versus the benefits of correct decisions (ID the guilty suspect, reject a lineup that contains an innocent suspect). Most policymakers agree that it is more important to protect the innocent (limit IDs of innocent suspects) than it is to implicate the guilty. Blackstone (1769, p. 352) famously said that it is "... better that ten guilty persons escape than that one innocent suffer." If the cost of a false ID is 10× greater than that of a miss, eyewitnesses should set a conservative criterion, and unbiased instructions help accomplish that. However, it is important to point out that the choice of these utilities is a matter for society and policymakers, not for eyewitness researchers. Once the

utilities are agreed upon, signal detection theory provides the machinery for converting the utilities, given the base rates of guilty and innocent suspects being placed into lineups, into an optimal criterion placement (for a review of these issues, see [Clark et al., 2015](#)). Base rates are notoriously difficult to estimate, however. Wixted et al. (2016) analyzed the Houston Police Department field data and concluded, based on a signal-detection model, that 35% of the lineups contained a guilty suspect and 65% contained an innocent suspect. To date, this is the only principled estimate of the base rate in one jurisdiction in the US. Therefore, more field data from other countries, jurisdictions and time-points are required before utilities can be converted into recommending an optimal criterion placement.

2.29.5.4.1.2 Adding a "Do Not Know" Option

Another instruction that can be offered to eyewitnesses is the option of reporting that they “do not know”, which is another recommendation included in the 2020 AP-LS White paper (Wells et al., 2020). [Weber and Perfect \(2012\)](#) conducted a study that examined the inclusion of a “do not know” option. Participants viewed a mock crime video, a distracting video, and then a target-present or target-absent showup. Some participants were required to make a “yes” or “no” decision about the face in the showup, other participants were allowed to choose “do not know” if they desired. Collapsed over retention interval (which was 3 min or 3 weeks), those participants with the option to choose “do not know” made better discriminations ($d' = .77$) than those

participants without that option ($d' = .28$, response criterion position was more conservative, $c = 0.75$ vs. $c = 0.47$, respectively). Other studies have also replicated the discriminability advantage on lineup tasks (Brewer et al., 2010; Perfect and Weber, 2012; Varnish et al., in prep). Steblay and Philips (2011), however, found the opposite pattern of results (see Varnish et al., in prep, for a review).

Considering reliability, Varnish et al. (in prep) recently conducted an experiment and high-confidence identifications were similarly reliable across conditions where there was, versus was not, a “do not know” option. In this study, although discriminability was descriptively higher when there was the option to choose “do not know”, this was not statistically significant in an ROC analysis. Therefore, it is perhaps unsurprising that reliability was equally good in both conditions. More research needs to be conducted to examine the effects of an explicit ‘do not know’ option on reliability.

2.29.5.4.1.3 Appearance Change Instructions

Researchers also have explored the impact of instructing eyewitnesses that a perpetrator may have changed his appearance since committing a crime. Charman and Wells (2007) had participants view a video with four culprits; this was followed by four lineups (two target-present and two target-absent, unbiased instructions). Half the participants received appearance change instructions and half did not. We computed d' and c from their data for each culprit and then took the average across culprits. We found that the appearance change instructions made participants slightly more liberal and slightly decreased discriminability. Molinaro et al. (2013) conducted a similar study and explicitly varied the amount of appearance change. They found that the appearance change instruction harmed discriminability and made responding more liberal, and

that the effects were similar across targets that differed in the amount of appearance change. Finally, Porter et al. (2014) had participants view a mock crime and used more extensive appearance change instructions. Experiment 2 included an explicit appearance change (addition of facial hair) for the target. Like the prior experiments, d' was lower for the appearance change condition (especially for the White culprit), although there was no response bias change. Overall, the data suggest that the appearance change instruction slightly decreases discriminability and tends to make participants' response biases more liberal. Porter et al. reported that participants receiving the appearance change instructions were less confident in their choices; but an assessment of the reliability of eyewitnesses receiving this instruction awaits further research.

2.29.5.4.2 Juror Instructions

Research has also been conducted on how jury instructions affect the evaluation of eyewitness evidence. Telfaire (United States v. Telfaire, 1972) instructions direct jurors to consider factors that could impact the accuracy of an eyewitness (e.g., viewing conditions) but fail to explain how these factors impact accuracy. Greene (1988) revised the Telfaire instructions to simplify them, and added information about how various factors (i.e., the Biggers criteria) impact eyewitness accuracy. But compared to control participants that received no cautionary instructions, the Telfaire participants were no better at distinguishing between good and poor eyewitnesses. The revised instructions did, however, make mock jurors more skeptical overall.

Similarly, New Jersey (2012) adopted an expanded, carefully designed (Schacter and Loftus, 2013) set of jury instructions. The goal of these instructions was to inform jurors about the current science of eyewitness memory, and how to use that knowledge to assess eyewitness testimony. Papailiou et al. (2015) assessed the effectiveness of these instructions. Mock jurors

viewed a 35-minute murder trial, which included either weak or strong eyewitness evidence. One half of each group was administered the New Jersey instructions, the other half was administered standard instructions. Jurors voted to convict about 25% of time given the standard instructions versus about 10% of the time given the enhanced instructions. More importantly, neither set of instructions aided mock jurors' ability to distinguish between the weak and strong eyewitness evidence (despite clear differences in evidence quality, Papailiou et al., 2015).

In sum, instructions to eyewitnesses or jurors appear to make both more conservative. If the goal is to reduce false IDs of innocent suspects, this is commendable. But if the goal of “better” instructions is to enhance discriminability or the reliability of eyewitness evidence, it would appear that research efforts are better directed elsewhere. Moreover, if instructional variations are simply moving eyewitnesses' response criteria around, an alternative perspective is to consider the response confidence reported by eyewitnesses. According to signal detection theory, changes in response confidence are a direct reflection of the underlying response criteria (Wixted et al., 2015a). In other words, instead of conducting more research on the instructions administered to eyewitnesses or jurors, more research should be conducted on how best to assess eyewitness confidence, which, it appears, has the potential to significantly enhance the ability to distinguish between weak and strong eyewitness evidence, if jurors can be made to rely on the initial, fairly assessed, identification evidence.

2.29.6 Conclusion

This chapter has covered a host of variables that affect eyewitness memory and introduced a new classification system for those variables. Going forward with research on eyewitness identification, it is important to separately measure discriminability (e.g., with ROC analysis)

and reliability (with CAC analysis). Moreover, there is a need (see also Clark, 2008; Clark and Gronlund, 2015; Gronlund et al., 2015) to develop and test competing theories of discriminability and reliability, because eyewitness identification research generally has been guided by verbally specified theories or intuition, which makes it difficult to extract definitive predictions (Bjork, 1973; Lewandowsky, 1993) and slows cumulative progress. A formally specified model, on the other hand, forces a theoretician to be explicit about a model's assumptions, which makes predictions transparent and provides a check on reasoning biases (Hintzman, 1991).

Some progress has been made developing formally specified explanations. One such theory, WITNESS, was proposed by Clark (2003). WITNESS is a direct-access matching model (see Clark and Gronlund, 1996) with a signal-detection foundation. The model parameters are closely tied to the components relevant to eyewitness identification. The model has been used to explore filler selection, simultaneous and sequential lineups (Goodsell et al., 2010), and relative and absolute judgments (Clark et al., 2011; Fife et al., 2014). More recently, signal-detection based models of lineup memory have been proposed, which specify different decision rules that could underlie witness decision-making (Wixted et al., 2018). The Independent Observations model is the simplest (as described in Parametric Measure of Discriminability: d') and assumes that witnesses identify the most familiar face in the lineup if that face exceeds their decision-criterion. The Ensemble model, on the other hand, is a mathematical instantiation of the diagnostic feature-detection theory (Wixted & Mickes, 2014 described in the Simultaneous Versus Sequential Lineups section) and assumes that witnesses compare the most familiar face in the lineup to the average of the other faces, and then identify the most familiar face if the difference exceeds their decision-criterion. While these two models often make qualitatively

similar (and correct) predictions about the effects variables will have on eyewitness discriminability, recent research provides evidence in favor of the Ensemble model (e.g., Shen et al., 2023; Wixted et al., 2018). Research should continue to pursue the development and testing of formal models.

Moreover, we have focused here on the reliability of eyewitness identifications made by adults, but recent research suggests that the usefulness of identifications and memory evidence from other populations may have also been previously underestimated. For example, previous eyewitness research with children using correlation or calibration analyses reported that children who are younger than 12 have not yet fully developed the skills to monitor their memory, or to use confidence scales to indicate accuracy (e.g., Brewer & Day, 2005; Keast et al., 2007; Parker & Carranza, 1989; Parker & Ryan, 1993; Powell et al., 2013) and so concluded that confidence is not a useful guide to accuracy for children's identification responses. Critically, this conclusion has informed legal guidance worldwide. Powell et al. (2013), for example, has been cited by superior courts in every jurisdiction in Australia and New Zealand. Yet, using CAC analysis, subsequent eyewitness research and reanalyses of basic recognition memory tasks, where children learn and are tested on lists of items, show that confidence provides considerable information about memory accuracy from at least age 8, but possibly younger depending on the task (Winsor et al., 2021). On a lineup task (arguably the most complex task reviewed by the authors), Winsor et al. (2021) found that 10-12 year olds were ~75% accurate at low confidence and 97% accurate at high confidence, 7-9 year olds were ~72% accurate at low confidence and 91% accurate at high confidence, while 4-6 year olds did not show a confidence accuracy relationship and were ~83% accurate at low confidence and 87% accurate at high confidence (see also Bruer et al., 2017). Moreover, confidence appears to be more informative than other

indicators commonly used to infer memory accuracy, such as the child's age, or their spontaneous verbal utterances (e.g., "umm") or body gestures (e.g., shrugs) of uncertainty (Hayre et al., 2023; Ingham et al., 2020). This suggests that identifications made by children can also be reliable when appropriate confidence measures are used to estimate their likely memory accuracy. Future research should continue to investigate other populations of witnesses whose memories are tested in the Criminal Justice System.

To conclude, many may find surprising what we have argued—that eyewitness identifications can be trustworthy, in the proper circumstances. The widely held alternative view that eyewitnesses are always unreliable, arose because of the indisputable evidence regarding the malleability of memory, coupled with the large number of wrongful convictions due to faulty eyewitness IDs. But measurement errors involving how to assess discriminability (ROC analysis, not diagnosticity) and how to assess the relationship between confidence and accuracy (calibration or CAC analysis, not the point-biserial correlation) contributed to the fact that researchers reached some premature conclusions (for a review, see [Gronlund et al., 2015](#)). But the largest error, arguably, was to recommend discounting eyewitness confidence. Our review indicates that eyewitnesses often can calibrate their confidence to reflect the likelihood that they are making a correct decision or an error, because an identification made with high confidence is, in many circumstances examined to date, much more likely to be an identification of a perpetrator. Conversely, an eyewitness who makes an identification with low confidence is likely indicating that an identification may not be trustworthy. Although this chapter has not emphasized the low confidence end of the CAC plot, the implications of low-confidence IDs also warrant careful consideration. Although a high-confidence ID is likely to signal a guilty suspect,

a low-confidence ID should lead the police to consider that their suspect is innocent. By ignoring confidence, the criminal justice system has missed the opportunity to utilize information that can be of great value.

This is most painfully evident by considering the typical profile of a DNA exoneration case involving eyewitness misidentification, where the initial identifications often were made with low confidence (at best) that was later inflated into a highly confident identification (see [Garrett, 2011](#)). By ignoring the initial level of confidence expressed on a first, fair, test of memory, and instead presenting jurors with a highly confident witness in court, it is not surprising that the jurors in these cases reached the conclusion that the eyewitnesses were accurate in their misidentifications. In those cases, had initial confidence been made known and relied upon by the jurors, those wrongfully convicted individuals may not have suffered, and the actual perpetrators of these crimes could have been apprehended sooner.

References

Ackerman, J.M., Shapiro, J.R., Neuberg, S.L., Kenrick, D.T., Becker, D.V., Griskevicius, V., Maner, J.K., Schaller, M., 2006. They all look the same to me (unless they're angry): from out-group homogeneity to out-group heterogeneity. *Psychol. Sci.* 17, 836–840.

Alogna, V.K., Attaya, M.K., Aucoin, P., Bahník, S., Birch, S., Birt, A.R., Zwaan, R.A., 2014. Registered replication report: Schooler & Engstler-Schooler (1990). *Perspect. Psychol. Sci.* 9, 556–578.

Amendola, K.L., Wixted, J.T., 2015. Comparing the diagnostic accuracy of suspect identifications made by actual eyewitnesses from simultaneous and sequential lineups in a randomized field trial. *J. Exp. Criminol.* 11, 263–284.

Bartlett, F.C., 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

Behrman, B.W., Davey, S.L., 2001. Eyewitness identification in actual criminal cases: an archival analysis. *Law Hum. Behav.* 25, 475–491.

Bennett, T. C., Collins, W., Flowe, H.D., Canham, R & Colloff, M.F. in prep. How are identification parades constructed in the United Kingdom?: A survey of identification officers.

Berkowitz, S. R., Garrett, B. L., Fenn, K. M., & Loftus, E. F. 2022. Convicting with confidence? Why we should not over-rely on eyewitness confidence. *Memory*, 30(1), 10-15.

Bernstein, M.J., Young, S.G., Hugenberg, K., 2007. The cross-category effect: mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychol. Sci.* 18, 706–712.

Bjork, R.A., 1973. Why mathematical models? *Am. Psychol.* 28, 426–433.

Blackstone, W., 1769. *Book IV. Commentaries on the Laws of England*, vol. II. Duyckinck, Long, Collins & Hannay, and Collins & Co, New York, NY.

Bonham, A.J., González-Vallejo, C., 2009. Assessment of calibration for reconstructed eye-witness memories. *Acta Psychol.* 131, 34–52.

Bornstein, B.H., Hamm, J.A., 2012. Jury instruction on witness identification. *Court Rev.* 48, 48–53.

Bornstein, B.H., Laub, C.E., Meissner, C.A., Susa, K.J., 2013. The cross-race effect: resistant to instructions. *J. Criminol.* 2013, 6.

Bradfield Douglass, A., Steblay, N., 2006. Memory distortion in eyewitnesses: a meta-analysis of the post-identification feedback effect. *Appl. Cogn. Psychol.* 20, 859–869.

Brewer, N., Day, K. 2005. The confidence-accuracy and decision latency-accuracy relationships in children's eyewitness identification. *Psychiatr Psychol Law*, 12(1), 119–128.

Brewer, N., Keast, A., & Sauer, J. D. 2010. Children's eyewitness identification performance: Effects of a Not Sure response option and accuracy motivation. *Leg. Crim. Psychol.*, 15(2), 261-277.

Brewer, N., Wells, G.L., 2006. The confidence-accuracy relation in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *J. Exp. Psychol. Appl.* 12, 11–30.

Brigham, J.C., Malpass, R.S., 1985. The role of experience and contact in the recognition of faces of own- and other-race persons. *J. Soc. Issues* 41, 139–155.

Bruer, K. C., Fitzgerald, R. J., Price, H. L., Sauer, J. D. 2017. How sure are you that this is the man you saw? Child witnesses can use confidence judgments to identify a target. *L. & Hum. Behav.*, 41(6), 541–555.

Buchanan, T.W., Lovallo, W.R., 2001. Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology* 26, 307–317.

Carlson, C. A., Jones, A. R., Whittington, J. E., Lockamy, R. F., Carlson, M. A., & Wooten, A. R. 2019. Lineup fairness: Propitious heterogeneity and the diagnostic feature-detection hypothesis. *Cog. Res.: Princ. and Impl.*, 4(1), 1–16.

Carlson, C.A., Carlson, M.A., 2014. An evaluation of perpetrator distinctiveness, weapon presence, and lineup presentation using ROC analysis. *J. Appl. Res. Mem. Cogn.* 3, 45–53.

Carlson, C.A., Dias, J.L., Weatherford, D.R., Carlson, M.A., 2016. An investigation of the weapon focus effect and the confidence-accuracy relationship for eyewitness identification. *J. Appl. Res. Mem. Cogn.*, 6(1), 82–92.

Carlson, C.A., Gronlund, S.D., Clark, S.E., 2008. Lineup composition, suspect position, and the sequential lineup advantage. *J. Exp. Psychol. Appl.* 14, 118–128.

Chance, J.E., Goldstein, A.G., 1981. Depth of processing in response to own and other race faces. *Pers. Soc. Psychol. Bull.* 7, 475–480.

Charman, S.D., Wells, G.L., 2007. Eyewitness lineups: is the appearance-change instruction a good idea? *Law Hum. Behav.* 31, 3–22.

Chin, J.M., Schooler, J.W., 2008. Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *Eur. J. Cogn. Psychol.* 20, 396–413.

Christianson, S.A., Hübner, B., 1993. Hands up! A study of witnesses' emotional reactions and memories associated with bank robberies. *Appl. Cogn. Psychol.* 7, 365–379.

Clare, J., Lewandowsky, S., 2004. Verbalizing facial memory: criterion effects in verbal overshadowing. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 739–755.

Clark, S., 2012. Costs and benefits of eyewitness identification reform: psychological science and public policy. *Perspect. Psychol. Sci.* 7, 238–259.

Clark, S.E., 2003. A memory and decision model for eyewitness identification. *Appl. Cogn. Psychol.* 17, 629–654.

Clark, S.E., 2008. The importance (necessity) of computational modelling for eyewitness identification research. *Appl. Cogn. Psychol.* 22, 803–813.

Clark, S.E., Benjamin, A.S., Wixted, J.T., Mickes, L., Gronlund, S.D., 2015. Eyewitness identification and the accuracy of the criminal justice system. *Policy Insights Behav. Brain Sci.* 2, 175–186.

Clark, S.E., Brower, G., Rosenthal, R., Hicks, J.M., Moreland, M.B., 2013. Lineup administrator influences on eyewitness identification and confidence. *J. Appl. Res. Mem. Cogn.* 2, 158–165.

Clark, S.E., Erickson, M.A., Breneman, J., 2011. Probative value of absolute and relative judgments in eyewitness identification. *Law Hum. Behav.* 35, 364–380.

Clark, S.E., Gronlund, S.D., 1996. Global matching models of recognition memory: how the models match the data. *Psychon. Bull. Rev.* 3, 37–60.

Clark, S.E., Gronlund, S.D., 2015. Mathematical modeling shows that compelling stories do not make for accurate descriptions of data. In: Raaijmakers, J.G.W. , Goldstone, R. , Steyvers, M. ,

Criss, A. , Nosofsky, R.M. (Eds.), *Cognitive Modeling in Perception and Memory: A Festschrift for Richard M. Shiffrin*. Psychology Press, pp. 245–258.

Clark, S.E., Marshall, T.E., Rosenthal, R., 2009. Lineup administrator influences on eyewitness identification decisions. *J. Exp. Psychol. Appl.* 15, 63–75.

Clark, S.E., Moreland, M.B., Gronlund, S.D., 2014. Evolution of theory and data in eyewitness identification reform. *Psychon. Bull. Rev.* 21, 251–267.

Clark, S.E., Rush, R.A., Moreland, M.B., 2013. Constructing the lineup: law, reform, theory, and data. In: Cutler, B. (Ed.), *Reform of Eyewitness Identification Procedures*. APA Publications, Washington, DC.

Colloff, M. F., & Wixted, J. T. 2020. Why are lineups better than showups? A test of the filler siphoning and enhanced discriminability accounts. *J. Exp. Psychol.: Appl.*, 26(1), 124–143.

Colloff, M. F., Flowe, H. D., Smith, H. J., Seale-Carlisle, T. M., Meissner, C. A., Rockey, J. C., Pande, B., Kujur, P., Parveen, N., Chandel, P., Singh, M. M., Pradhan, & S., Parganiha, A. 2022. Active exploration of faces in police lineups increases discrimination accuracy. *Am. Psychol.*, 77(2), 196-220.

Colloff, M. F., Seale-Carlisle, T. M., Karagolu, N., Maltby, J., Smith, H. J., Smith, L., Maltby, J., Yaremenko, S., & Flowe, H. D. 2021. Perpetrator pose reinstatement during a lineup test increases discrimination accuracy. *Sci. Rep.*, 11:13830.

Colloff, M. F., Wade, K. A., & Strange, D. 2016. Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychol. Sci.*, 27, 1227–1239.

Colloff, M. F., Wade, K. A., Strange, D., & Wixted, J. T. 2018. Filler siphoning theory does not predict the effect of lineup fairness on the ability to discriminate innocent from guilty suspects: Reply to Smith, Wells, Smalarz, & Lampinen 2018. *Psychol. Sci.*, 29, 1552-1557.

Colloff, M. F., Wade, K. A., Wixted, J. T., & Maylor, E. A. 2017. A signal-detection analysis of eyewitness identification across the adult lifespan. *Psychol. Aging*, 32(3), 243–258.

Colloff, M. F., Wilson, B. M., Seale-Carlisle, T. M., & Wixted, J. T. 2021. Optimizing the selection of fillers in police lineups. *Proc. Natl. Acad. Sci.*, 118(8), e2017292118.

Cutler, B.L. & Fisher, R.P. 1990. Live lineups, videotaped lineups, and photoarrays. *Forensic Reports*, 3(4), 439–448.

Cutler, B.L., Penrod, S.D., Martens, T.K., 1987. Improving the reliability of eyewitness identification: putting context into context. *J. Appl. Psychol.* 72, 629–637.

Darling, S., Valentine, T., Memon, A., 2007. Selection of lineup foils in operational contexts. *Appl. Cogn. Psychol.* 22, 159–169.

Davies, G. M., Smith, S., Blincoe, C. 2008. A “weapon focus” effect in children. *Psychology, Crime & Law*, 14, 19–28.

Deffenbacher, K.A., Bornstein, B.H., Penrod, S.D., McGorty, E.K., 2004. A meta-analytic review of the effects of high stress on eyewitness memory. *Law Hum. Behav.* 28, 687–706.

DeGutis, J., Mercado, R.J., Wilmer, J., Rosenblatt, 2013. Individual differences in holistic processing predict the own-race advantage in recognition memory. *PLoS One*.

Devlin, Lord P. 1976. Report to the Secretary of State for the Home Department on the Departmental Committee on Evidence of Identification in Criminal Cases. HMSO.

Dobolyi, D.G., Dodson, C.S., 2013. Eyewitness confidence in simultaneous and sequential lineups: a criterion shift account for sequential mistaken identification overconfidence. *J. Exp. Psychol. Appl.* 19, 345–357.

Ebbesen, E.B., Flowe, H.D., 2002. Simultaneous v. Sequential Lineups: What Do We Really Know? Retrieved from:

<http://www2.le.ac.uk/departments/psychology/ppl/hf49/SimSeq%20Submit.pdf>.

Fife, D., Perry, C., Gronlund, S.D., 2014. Revisiting absolute and relative judgments in the witness model. *Psychon. Bull. Rev.* 21, 479–487.

Fitzgerald, R. J., Price, H. L., & Valentine, T. 2018. Eyewitness identification: Live, photo, and video lineups. *Psychol., Pub. Pol., and Law*, 24(3), 307–325.

Fitzgerald, R. J., Rubínová, E., & Juncu, S. 2021. Eyewitness identification around the world. In Smith, A. M., Toglia, M., & Lampinen, J. M. (Eds.), *Meth., Meas., and Theories in Eyewitness Identification Tasks*. Taylor and Francis.

Freeman, J.B., Pauker, K., Sanchez, D.T., 2016. A perceptual pathway to bias: interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychol. Sci.* 27, 502–517.

Garrett, B., 2011. *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*. Harvard University Press, Cambridge, MA.

Golby, A.J., Gabrieli, J.D.E., Chiao, J.Y., Eberhardt, J.L., 2001. Differential responses in the fusiform region to same-race and other-race faces. *Nat. Neurosci.* 4, 845–850.

Goodsell, C.A., Gronlund, S.D., Carlson, C.A., 2010. Exploring the sequential lineup advantage using WITNESS. *Law Hum. Behav.* 34, 445–459.

Goodsell, C.A., Wetmore, S.A., Neuschatz, J.S., Gronlund, S.D., 2013. Showups vs. lineups: a review of two identification techniques. In: Cutler, B. (Ed.), *Reform of Eyewitness Identification Procedures*. APA Publications.

Greathouse, S.M., Kovera, M.B., 2009. Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law Hum. Behav.* 33, 70–82.

Green, D.M., Swets, J.A., 1966. *Signal Detection Theory and Psychophysics*. Wiley, New York.

Greene, E., 1988. Judge's instruction on eyewitness testimony: evaluation and revision. *J. Appl. Soc. Psychol.* 18, 353–376.

Gronlund, S.D., Carlson, C.A., Dailey, S.B., Goodsell, C.A., 2009. Robustness of the sequential lineup advantage. *J. Exp. Psychol. Appl.* 15, 140–152.

Gronlund, S.D., Carlson, C.A., Neuschatz, J.S., Goodsell, C.A., Wetmore, S., Wooten, A., et al., 2012. Showups versus lineups: an evaluation using ROC analysis. *J. Appl. Res. Mem. Cogn.* 1, 221–228.

Gronlund, S.D., Mickes, L., Wixted, J.T., Clark, S.E., 2015. Conducting an eyewitness lineup: how the research got it wrong. In: Ross, B.H. (Ed.), *The Psychology of Learning and Motivation*, vol. 63. Academic Press, Waltham, MA, pp. 1–43.

Gronlund, S.D., Wixted, J.T., Mickes, L., 2014. Evaluating eyewitness identification procedures using ROC analysis. *Curr. Dir. Psychol. Sci.* 23, 3–10.

Hasel, L.E., Kassin, S.M., 2009. On the presumption of evidentiary independence: can confessions corrupt eyewitness identifications? *Psychol. Sci.* 20, 122–126.

Haw, R.M., Fisher, R.P., 2004. Effects of administrator-witness contact on eyewitness identification accuracy. *J. Appl. Psychol.* 89, 1106–1112.

Hayre, R., Ingham, M., Smith, S., Findel, B., Sargent, C., Colloff, M. F. 2023. Metacognitive Monitoring of Memory Performance Using Implicit & Explicit Measures in 5- to 11-year-olds. Annual Meeting of the Psychonomic Society, San Francisco, California, US.

Hintzman, D.L., 1991. Why are formal models useful in psychology? In: Hockley, W.E. , Lewandowsky, S. (Eds.), *Relating Theory and Data: Essays on Human Memory in Honor of Bennet B. Murdock*. Erlbaum, Hillsdale, NJ, pp. 39–56.

Holdstock, J. S., Dalton, P., May, K., Boogert, S. & Mickes, L. 2022. Lineup identification in young and older witnesses: Does describing the criminal help or hinder?" *Cognitive Research: Principles and Implications*, 7(51).

Hope, L., & Wright, D. 2007. Beyond unusual? Examining the role of attention in the weapon focus effect. *Applied Cognitive Psychology*, 21, 951–961.

Hoscheidt, S.M., LaBar, K.S., Ryan, L., Jacobs, W.J., Nadel, L., 2014. Encoding negative events under stress: high subjective arousal is related to accurate emotional memory despite misinformation exposure. *Neurobiol. Learn. Mem.* 112, 237–247.

Houriham, K.L., Benjamin, A.S., Lui, X., 2012. A cross-race effect in metamemory: predictions of face recognition are more accurate for members of our own race. *J. Appl. Res. Mem. Cogn.* 1, 158–162.

Hugenberg, K., Young, S.G., Bernstein, M.J., Sacco, D.F., 2010. The categorization-individuation model: an integrative account of the other-race recognition deficit. *Psychol. Rev.* 117, 1168–1187.

Hulse, L.M., Memon, A., 2006. Fatal impact? The effects of emotional arousal and weapon presence on police officers' memories for a simulated crime. *Leg. Criminol. Psychol.* 11, 313–325.

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

Ingham, M. P., Hayre, R. K., Rockowitz, S. R., Elsherit, M., Colloff, M. F. 2020. Metacognitive Measures as Predictors of Episodic Memory Accuracy in Children. Annual Meeting of the Psychonomic Society, Virtual event.

Innocence Project, 2009. As More States Weigh Improving Lineups, New Innocence Project Report Shows Extent of the Problem and Effectiveness of Reform.

<http://www.innocenceproject.org/as-more-states-weigh-improving-lineups-new-innocence-project-report-shows-extent-of-the-problem-and-effectiveness-of-reform/>.

Innocence Project, 2015. Innocence Project Understand the Causes: the Causes of Wrongful Conviction. Innocence Project, New York. <http://www.innocenceproject.org/causes-wrongful-conviction>.

Johnson, K.J., Fredrickson, B.L., 2005. “We all look the same to me” Positive emotions eliminate the own-race bias in face recognition. *Psychol. Sci.* 16, 875–881.

Johnson, M.K., Hashtroudi, S., Lindsay, D.S., 1993. Source monitoring. *Psychol. Bull.* 114, 3–28.

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. *J. Exp. Psychol.: Learn., Mem., and Cogn.*, 22(5), 1304–1316.

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. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1304–1316.

Kaesler, M., Dunn, J. C., & Semmler, C. under review. Clarifying the Effects of Sequential Item Presentation in the Police Lineup Task.

Kaesler, M., Dunn, J. C., Ransom, K., & Semmler, C. 2020. Do sequential lineups impair discriminability? *Cog. Res.: Princ. and Impl.*, 5(1), 35.

Keast, A., Brewer, N., Wells, G. L. 2007. Children’s metacognitive judgments in an eyewitness identification task. *J. Exp. Child Psychol.*, 97(4), 286–314.

Kelly, L., Lovett, J. & Regan, L. 2005. A Gap or a Chasm? Attrition in Reported Rape Cases, *Home Off. Res. Stud.*, 293, Home Office.

Kirschbaum, C., Pirke, K.M., Hellhammer, D.H., 1993. The “Trier Social Stress Test”—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 76–81.

Kovera, M. B., & Evelo, A. J. 2017. The case for double-blind lineup administration. *Psychol., Pub. Pol., and Law*, 23(4), 421–437.

Lampinen, J.M., 2016. ROC analyses in eyewitness identification research. *J. Appl. Res. Mem. Cogn.* 5 (1), 21–33.

Lewandowsky, S., 1993. The rewards and hazards of computer simulations. *Psychol. Sci.* 4, 236–243.

Lindsay, D.S., Johnson, M.K., 1989. The eyewitness suggestibility effect and memory for source. *Mem. Cogn.* 17, 349–358.

Lindsay, R.C.L., 1999. Applying applied research: selling the sequential line-up. *Appl. Cogn. Psychol.* 13, 219–225.

Lindsay, R.C.L., Lea, J.A., Fulford, J.A., 1991. Sequential lineup presentation: technique matters. *J. Appl. Psychol.* 76, 741–745.

Lindsay, R.C.L., Lea, J.A., Nosworthy, G.J., Fulford, J.A., Hector, J., LeVan, V., Seabrook, C., 1991. Biased lineups: sequential presentation reduces the problem. *J. Appl. Psychol.* 76, 796–802.

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 Hum. Behav.* 18, 527–541.

Lindsay, R.C.L., Wells, G.L., 1985. Improving eyewitness identifications from lineups: simultaneous versus sequential lineup presentation. *J. Appl. Psychol.* 70, 556–564.

Loftus, E.F., 1979. Malleability of human memory. *Am. Sci.* 67, 312–320.

Loftus, E.F., 2005. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn. Mem.* 12, 361–366.

Loftus, E.F., Loftus, G.R., Messo, J., 1987. Some facts about “weapon focus”. *Law Hum. Behav.* 11, 55–62.

Loftus, E.F., Palmer, J.C., 1974. Reconstruction of auto-mobile destruction: an example of the interaction between language and memory. *J. Verbal Learn. Verbal Behav.* 13, 585–589.

Lusted, L.B., 1971. Decision-making studies in patient management. *N. Engl. J. Med.* 284, 416–424.

Lusted, L.B., 1971. Signal detectability and medical decision-making. *Science* 171, 1217–1219.

Luus, C.A.E., Wells, G.L., 1991. Eyewitness identification and the selection of distracters for lineups. *Law Hum. Behav.* 15, 43–57.

Maass, A., Kohnken, G., 1989. Eyewitness identification: simulating the “weapon effect”. *Law Hum. Behav.* 13, 397–408.

Macmillan, N.A., Creelman, C.D., 2005. *Detection Theory: A User’s Guide*, second ed. Erlbaum, Mahwah, NJ.

Mahmood, A., Roughton, G. L., & Colloff, M. F. 2022. Replicating similar, but non-identical features in lineups improves eyewitness accuracy. Poster presented at the Annual Meeting of the Psychonomic Society, Boston and online (hybrid), 2022.

Malpass, R.S., Devine, P.G., 1981. Eyewitness identification: lineup instructions and the absence of the offender. *J. Appl. Psychol.* 66, 482–489.

Malpass, R.S., Kravitz, J., 1969. Recognition for faces of own and other race. *J. Pers. Soc. Psychol.* 13, 330–334.

Marmurek, H. H., Rusyn, R., Zgardau, A., & Zgardau, A. M. 2022. Verbal overshadowing at an immediate Task-Test delay is independent of Video-Task delay. *J. Cog. Psychol.*, 34(2), 243-248.

Marr, C., Quaedflieg, C.W.E.M., Otgaar, H., Hope, L., & Sauerland, M. 2021. Facing stress: No effect of acute stress at encoding or retrieval on face recognition memory. *Acta Psychol.*, 219:103376.

Meissner, C.A., Brigham, J.C., 2001. A meta-analysis of the verbal overshadowing effect in face identification. *Appl. Cogn. Psychol.* 15, 603–616.

Meissner, C.A., Tredoux, C.G., Parker, J.F., MacLin, O.H., 2005. Eyewitness decisions in simultaneous and sequential lineups: a dual process signal detection theory analysis. *Mem. Cogn.* 33, 783–792.

Meyer, M., Colloff, M. F., Bennett, T. C., Hirata, E., Kohl, A., Stevens, L. M., Smith, H. M. J., Staudigl, T., & Flowe, H. D. 2023. Enabling witnesses to actively explore faces and reinstate study-test pose during a lineup increases discriminability. *Proc. Natl. Acad. Sci.*, 120(41), e2301845120.

Mickes, L., 2015. Receiver operating characteristic analysis and confidence accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *J. Appl. Res. Mem. Cogn.* 4, 93–102.

Mickes, L., 2016. The effects of verbal descriptions on eyewitness memory: implications for the real-world. *J. Appl. Res. Mem. Cogn.* 5, 270–276.

Mickes, L., Flowe, H.D., Wixted, J.T., 2012. Receiver operating characteristic analysis of eyewitness memory: comparing the diagnostic accuracy of simultaneous versus sequential lineups. *J. Exp. Psychol. Appl.* 18 (4), 361–376.

Mickes, L., Moreland, M.B., Clark, S.E., Wixted, J.T., 2014. Missing the information needed to perform ROC analysis? Then compute d' , not the diagnosticity ratio. *J. Appl. Res. Mem. Cogn.* 3, 58–62.

Mickes, L., Seale-Carlisle, T. M., Wetmore, S. A., Gronlund, S. D., Clark, S. E., Carlson, C. A., Goodsell, C. A., Weatherford, D., & Wixted, J. T. 2017. ROCs in Eyewitness Identification: Instructions versus Confidence Ratings. *Appl. Cogn. Psychol.*, 31(5), 467-477.

Mickes, L., Wixted, J.T., 2015. On the applied implications of the “verbal overshadowing effect”. *Perspect. Psychol. Sci.* 10, 400–403.

Molinaro, P.F., Arndorfer, A., Charman, S.D., 2013. Appearance-change instructions effects on eyewitness lineup identification accuracy are not moderated by the amount of appearance change. *Law Hum. Behav.* 37, 432–440.

Morgan D. P., Tamminen J., Seale-Carlisle T.M., & Mickes, L. 2019. The impact of sleep on eyewitness identifications. *Royal Society Open Science*. Registered report.

Morgan III, C.A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., et al., 2004.

Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *Int. J. Law Psychiatry* 27, 265–279.

Morgan III, C.A., Southwick, S., Steffian, G., Hazlett, G.A., Loftus, E.F., 2013. Misinformation can influence memory for recently experienced, highly stressful events. *Int. J. Law Psychiatry* 36, 11–17.

National Research Council, 2014. *Identifying the Culprit: Assessing Eyewitness Identification*. .

Ng, W., Lindsay, R.C.L., 1994. Cross-race facial recognition: failure of the contact hypothesis. *J. Cross Cult. Psychol.* 25, 217–232.

Nguyen, T. B., Pezdek, K., & Wixted, J. T. 2017. Evidence for a confidence-accuracy relationship in memory for same- and cross-race faces. *Quart. J. Exp. Psychol.: Hum. Exp. Psychol.*, 70(12), 2518–2534.

O'Rourke, T., Penrod, S., Curler, B., Stuve, T., 1989. The external validity of eyewitness identification research: generalizing across subject populations. *Law Hum. Behav.* 13, 385–395.

Open Science Collaboration., 2015. *Science* 349, aac4716.

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. *J. Exp. Psychol.: Appl.*, 19(1), 55–71.

Palmer, M., Brewer, N., Weber, N., Nagesh, A., 2013. The confidence-accuracy relationship for eyewitness identification decisions: effects of exposure duration, retention interval, and divided attention. *J. Exp. Psychol. Appl.* 19, 55–71.

Palmer, M.A., Brewer, N., 2012. Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law Hum. Behav.* 36, 247–255.

- Papailiou, A.P., Yokum, D.V., Robertson, C.T., 2015. The novel New Jersey eyewitness instruction induces skepticism but not sensitivity. *PLoS One*. 10.
- Papesh, M.H., Goldinger, S.D., 2010. A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition* 116, 283–288.
- Parker, J. F., Carranza, L. E. 1989. Eyewitness testimony of children in target-present and target-absent lineups. *L. & Hum. Behav.*,13(2), 133–149.
- Parker, J. F., Ryan, V. 1993. An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. *L. & Hum. Behav.*, 17(1), 11–26.
- Pashler, H., Wagenmakers, E.-J., 2012. Editor's introduction to the special section on replicability in psychology science: a crisis of confidence? *Perspect. Psychol. Sci.* 7, 528–530.
- Paterson, H.M., Kemp, R.I., Ng, J.R., 2011. Combating co-witness contamination: attempting to decrease the negative effects of discussion on eyewitness memory. *Appl. Cogn. Psychol.* 25, 43–52.
- Payne, J.D., Jackson, E.D., Hoscheidt, S., Ryan, L., Jacobs, W.J., Nadel, L., 2007. Stress administered prior to encoding impairs neutral but enhances emotional long-term episodic memories. *Learn. Mem.* 14, 861–868.
- Payne, J.D., Nadel, L., Allen, J.J., Thomas, K.G., Jacobs, W.J., 2002. The effects of experimentally induced stress on false recognition. *Memory* 10, 1–6.
- Penrod, S., Cutler, B., 1995. Witness confidence and witness accuracy: assessing their forensic relation. *Psychol. Public Policy Law* 4, 817–845.

Perfect, T. J., & Weber, N. 2012. How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back?. *J. Exp. Psychol.: Learn., Mem., and Cogn.*, 38(6), 1810-1818.

Pickel, K.L., 1998. Unusualness and threat as possible causes of “weapon focus”. *Memory* 6, 277–295.

Platz, S.J., Hosch, H.M., 1988. Cross-racial ethnic eyewitness identification: a field study. *J. Appl. Soc. Psychol.* 18, 972–984.

Police and Criminal Evidence Act 1984. Codes of Practice, Code D 2017.

Police Executive Research Forum, 2013. A National Survey of Eyewitness Identification Procedures in Law Enforcement Agencies. <http://www.policeforum.org/>.

Porter, D., Moss, A., Reisberg, D., 2014. The appearance-change instruction does not improve lineup identification accuracy. *Appl. Cogn. Psychol.* 28, 151–160.

Powell, M. B., Garry, M., Brewer, N. 2013. Eyewitness testimony. In I. Freckelton & H. Selby (Eds.), *Expert evidence: Law, practice, procedure and advocacy* (5th ed.). Thomson Reuters.

Read, J. D., Lindsay, D. S., & Nichols, T. 1998. The relation between confidence and accuracy in eyewitness identification studies: Is the conclusion changing? In C. P. Thomson, D. Bruce, J. D. Read, D. Hermann, D. Payne, & M. P. Toglia (Eds.), *Eyewitness Memory: Theoretical and Applied Perspectives*, pp. 107–130. Lawrence Erlbaum.

Roediger, H.L., McDermott, K.B., 1995. Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814.

Rosenthal, R., 2002. Covert communication in classrooms, clinics, courtrooms, and cubicles. *Am. Psychol.* 57, 839–849.

Rotello, C. M., & Chen, T. 2016. ROC curve analyses of eyewitness identification decisions: An analysis of the recent debate. *Cog. Res.: Princ. and Impl.*, 1(1), 1-12.

Rotello, C.M., Heit, E., Dube, C., 2015. When more data steer us wrong: replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychon. Bull. Rev.* 22, 944–954.

Roughton, G., Bennett, T., Flowe, H.D., Varnish. C., & Colloff, M.F. 2023. A Systematic Review of Lineup Filler Similarity. Presented at SARMAC XIV, International Conference, Nagoya, Japan, 2023.

Rubínová, E., Fitzgerald, R. J., Juncu, S., Ribbers, E., Hope, L., & Sauer, J. D. 2021. Live presentation for eyewitness identification is not superior to photo or video presentation. *J. Appl. Res. in Mem. and Cogn.*, 10(1), 167-176.

Russano, M.B., Dickinson, J.J., Greathouse, S.M., Kovera, M.B., 2006. “Why don't you take another look at number three?” Investigator knowledge and its effects on eyewitness confidence and identification decisions. *Cardozo Public Law Policy Ethics J.* 4, 355–379.

Sauer, J. D., Palmer, M. A., & Brewer, N. 2019. Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychol., Pub. Pol., and Law*, 25(3), 147–165.

Sauer, J., Brewer, N., Zweck, T., & Weber, N. 2010. The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law and Hum. Behav.*, 34(4), 337–347.

- Sauer, J., Brewer, N., Zweck, T., Weber, N., 2010. The effect of retention interval on the confidence-accuracy relationship for eyewitness identification. *Law Hum. Behav.* 34, 337–347.
- Sauerland, M., Raymaecker, L.H.C., Otgaar, H., Memon, A., Waltjen, T.T., Nivo, M., Slegers, C., Broers, N.J., Smeets, T., 2016. Stress, stress-induced cortisol responses, and eyewitness identification performance. *Behav. Sci. Law* 34, 580–594.
- Schacter, D.L., Loftus, E.L., 2013. Memory and law: what can cognitive neuroscience contribute? *Nat. Neurosci.* 16, 119–123.
- Schooler, J.W., 2002. Verbalization produces a transfer inappropriate processing shift. *Appl. Cogn. Psychol.* 16, 989–997.
- Schooler, J.W., Engstler-Schooler, T.Y., 1990. Verbal overshadowing of visual memories: some things are better left unsaid. *Cogn. Psychol.* 22, 36–71.
- Seale-Carlisle, T. M., Wetmore, S. A., Flowe, H. D., & Mickes, L. 2019. Designing police lineups to maximize memory performance. *J. Exp. Psychol.: Appl.*, 25(3), 410.
- Seale-Carlisle, T.M., & Mickes, L. 2016. US line-ups outperform UK line-ups. *R. Soc. open sci.*, 3: 160300.
- Shaw, J.I., Skolnick, P., 2001. Sex difference, weapon focus, and eyewitness reliability. *J. Soc. Psychol.* 134, 413–420.
- Shen, K. J., Colloff, M. F., Vul, E., Wilson, B. M., & Wixted, J. T. 2023. Modeling face similarity in police lineups. *Psychol. Rev.* 130(2), 432–461.
- Shriver, E.R., Hugenberg, K., 2010. Power, individuation, and the cross-race recognition deficit. *J. Exp. Soc. Psychol.* 46, 767–774.

Spearing, E. R., & Wade, K. A. 2022. Providing eyewitness confidence judgments during versus after eyewitness interviews does not affect the confidence–accuracy relationship. *J. Appl. Res. in Mem. and Cogn.* 11(1), 54–65.

Sporer, S. L. 2001. Recognizing faces of other ethnic groups: An integration of theories. *Psychol., Pub. Pol., and Law.* 7(1), 36–97.

Sporer, S.L., Penrod, S., Read, D., Cutler, B., 1995. Choosing, confidence, and accuracy: a meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychol. Bull.* 118, 315–327.

Starr, S.J., Metz, C.E., Lusted, L.B., Goodenough, D.J., 1975. Visual detection and localization of radiographic images. *Radiology* 116, 538–553.

Steblay, N., 1997. Social influence in eyewitness recall: a meta-analytic review of lineup instruction effects. *Law Hum. Behav.* 21, 283–397.

Steblay, N., Dysart, J., Fulero, S., Lindsay, R.C., 2001. Eyewitness accuracy rates in sequential and simultaneous lineup presentations: a meta-analytic comparison. *Law Hum. Behav.* 205, 459–473.

Steblay, N., Dysart, J., Fulero, S., Lindsay, R.C., 2003. Eyewitness accuracy rates in police showup and lineup presentations: a meta-analytic comparison. *Law Hum. Behav.* 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. *Psychol. Public Policy Law* 17 (1), 99–139.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.

Swets, J.A., Dawes, R.M., Monahan, J., 2000. Psychological science can improve diagnostic decisions. *Psychol. Sci. Public Interest* 1, 1–26.

Technical Working Group for Eyewitness Evidence, 1999. *Eyewitness Evidence: A Guide for Law Enforcement* [Booklet]. United States Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC.

The Christian Science Monitor, February 15, 2016. How Ohio Diners Helped Police Find Machete-wielding Attacker. <http://www.csmonitor.com/USA/Justice/2016/0212/How-Ohio-diners-helped-police-find-machete-wielding-attacker>.

Tulving, E. & Thomson, D.M. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.*, 80(5), 352–373.

Tunnicliff, J.L., Clark, S.E., 2000. Selecting foils for identification lineups: matching suspects or descriptions. *Law Hum. Behav.* 24, 231–258.

Tyler, T.R., 2003. Procedural justice, legitimacy, and the effective rule of law. *Crime Justice* 30, 283–357.

United States v. Telfaire, 469F.2d 552 (D. C. Cir. 1972). *Washington Pattern Instructions – Criminal* (1977). St. Paul: West Publishing Co.

Valentine, T., 1991. A unified account of the effects of distinctiveness, inversion and race on face recognition. *Q. J. Exp. Psychol.* 43, 161–204.

Valentine, T., Darling, S. & Memon, A. 2007. Do strict rules and moving images increase the reliability of sequential identification procedures? *Appl. Cog. Psychol.*, 21(7), 933–949.

Valentine, T., Mesout, J., 2008. Eyewitness identification under stress in the London Dungeon. *Appl. Cogn. Psychol.* 23, 151–161.

Valentine, T., Pickering, A., Darling, S., 2002. Characteristics of eyewitness identification that predict the outcome of real lineups. *Appl. Cogn. Psychol.* 17, 969–993.

Varnish, C., Condruz, & Colloff, M.F. in prep. An ROC and CAC analysis of lineup construction methods.

Wade, K. A., Nash, R. A., & Lindsay, D. S. 2018. Reasons to Doubt the Reliability of Eyewitness Memory: Commentary on Wixted, Mickes, and Fisher 2018. *Perspect. on Psychol. Sci.*, 13(3), 339–342.

Wagstaff, G.F., MacVeigh, J., Scott, L., Brunas-Wagstaff, J., Cole, J., 2003. Can laboratory findings on eyewitness testimony be generalized to the real world? An archival analysis of influence of violence, weapon presence, and age eyewitness accuracy. *J. Psychol.* 137, 17–28.

Weber, N., Brewer, N., 2004. Confidence-accuracy calibration in absolute and relative face recognition judgments. *J. Exp. Psychol. Appl.* 10, 156–172.

Weber, N., Perfect, T.J., 2012. Improving eyewitness identification accuracy by screening out those who say they don't know. *Law Hum. Behav.* 36, 28–36.

Wells G. L., Memon A., Penrod S. D. 2006. Eyewitness identification: Improving its probative value. *Psychol. Sci. in the Pub. Int.*, 7, 45–75.

Wells, G. L. 2001. Eyewitness lineups: Data, theory, and policy. *Psychol., Pub. Pol., and Law*, 7(4), 791–801.

- Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. 2020. Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Hum. Behav.*, 44(1), 3–36.
- Wells, G.L., 1978. Applied eyewitness-testimony research: system variables and estimator variables. *J. Pers. Soc. Psychol.* 12, 1546–1557.
- Wells, G.L., Bradfield, A.L., 1998. “Good, you identified the suspect”: feedback to eyewitnesses distorts their reports of the witnessing experience. *J. Appl. Psychol.* 83, 360–376.
- Wells, G.L., Lindsay, R.C., 1980. On estimating the diagnosticity of eyewitness nonidentifications. *Psychol. Bull.* 88, 776–784.
- Wells, G.L., Malpass, R.S., Lindsay, R.C.L., Fisher, R.P., Turtle, J.W., Fulero, S.M., 2000. From the lab to the police station: a successful application of eyewitness research. *Am. Psychol.* 55, 581–598.
- Wells, G.L., Murray, D.M., 1984. Eyewitness confidence. In: Wells, G.L., Loftus, E.F. (Eds.), *Eyewitness Testimony: Psychological Perspectives*. Cambridge University Press, New York, NY, pp. 155–170.
- Wells, G.L., Rydell, S.M., Seelau, E.P., 1993. The selection of distractors for eyewitness lineups. *J. Appl. Psychol.* 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. *J. Appl. Res. Mem. Cogn.* 4, 324–328.

Wells, G.L., Smith, A.M., Smalarz, L., 2015. ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *J. Appl. Res. Mem. Cogn.* 4, 324–328.

Wells, W., 2014. The Houston Police Department Eyewitness Identification Experiment: Analysis and Results. Retrieved from: <http://www.lemiltonline.org/research/projects.html>.

Wetmore, S.A., Neuschatz, J.S., Gronlund, S.D., Wooten, A., Goodsell, C.A., Carlson, C.A., 2015. Effect of retention interval on showup and lineup performance. *J. Appl. Res. Mem. Cogn.* 4, 8–14.

Wilson, B. M., & Colloff, M.F. 2020. Coherently creating full receiver operating characteristic curves of police lineups. Poster presented at the Annual Meeting of the Psychonomic Society, Online.

Wilson, B. M., Donnelly, K., Christenfeld, N., & Wixted, J. T. 2019. Making sense of sequential lineups: An experimental and theoretical analysis of position effects. *J. Mem. and Lang.*, 104, 108-125.

Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. 2018. The effects of verbal descriptions on performance in lineups and showups. *J. Exp. Psychol. Gen.*, 147(1), 113-124.

Winsor, A., Flowe, H. D., Seale-Carlisle, T. M., Killeen, I. M., Hett, D., Jores, T., Ingham, M., Lee, B., Stevens, L., Colloff, M. F. 2021. Children's expressions of certainty are informative. *J. Exp. Psychol.*, 150(11), 2387–2407.

Wise, R.A., Safer, M.A., Maro, C.M., 2011. What US law enforcement officers know and believe about eyewitness factors, eyewitness interviews and identification procedures. *Appl. Cogn. Psychol.* 25, 488–500.

Wixted, J. T., Mickes, L. 2018. Theoretical vs. empirical discriminability: the application of ROC methods to eyewitness identification. *Cogn. Res. Princ. Implic.* 3:9.

Wixted, J. T., Mickes, L. 2022. Eyewitness memory is reliable, but the criminal justice system is not. *Memory*, 30(1), 67-72.

Wixted, J. T., Wells, G. L. 2017. The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychol. Sci. Public Interest*, 18, 10-65.

Wixted, J.T., Mickes, L., 2012. The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspect. Psychol. Sci.* 7, 275–278.

Wixted, J.T., Mickes, L., 2014. A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychol. Rev.* 121, 262–276.

Wixted, J.T., Mickes, L., 2015. Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *J. Appl. Res. Mem. Cogn.* 4, 318–323.

Wixted, J.T., Mickes, L., 2015. ROC analysis measures objective discriminability for any eyewitness identification procedure. *J. Appl. Res. Mem. Cogn.* 4, 329–334.

Wixted, J.T., Mickes, L., 2022. Eyewitness memory is reliable, but the criminal justice system is not. *Memory*, 30, 67-72.

Wixted, J.T., Mickes, L., Clark, S.E., Gronlund, S.D., Roediger, H.L., 2015a. Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *Am. Psychol.* 70 (6), 515–526.

Wixted, J.T., Mickes, L., Dunn, J., Clark, S.E., Wells, W., 2015b. Estimating the reliability of eyewitness identifications from police lineups. *Proc. Natl. Acad. Sci. U.S.A.* 113, 304–309.

Wixted, J.T., Read, D., Lindsay, D.S., 2016. The effect of retention interval on the eyewitness identification confidence-accuracy relationship. *J. Appl. Res. Mem. Cogn.* 5 (2), 192-203.

Wogalter, M.S., Malpass, R.S., Mcquiston, D.E., 2004. A national survey of US police on preparation and conduct of identification lineups. *Psychol. Crime Law* 10, 69–82.

Wright, D.B., Boyd, C.E., Tredoux, C.G., 2001. A field study of own-race bias in South Africa and England. *Psychol. Public Policy Law* 7, 119–133.

Wright, D.B., Skagerberg, E.M., 2007. Postidentification feedback affects real eyewitnesses. *Psychol. Sci.* 18, 172–178.

Yates S. Q. 2017. Eyewitness identification: Procedures for conducting photo arrays. U.S. Department of Justice. Retrieved from <https://www.justice.gov/archives/opa/press-release/file/923201/download>

Zaragoza, M.S., Lane, S.M., 1994. Sources of misattribution and suggestibility of eyewitness testimony. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 934–945.

Zweig, M.H., Campbell, G., 1993. Receiver operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* 39, 561–577.

Figure 1 Six-person simultaneous lineup with five fillers and one suspect. If the suspect is the perpetrator the lineup is target-present; if the suspect is innocent the lineup is target-absent.

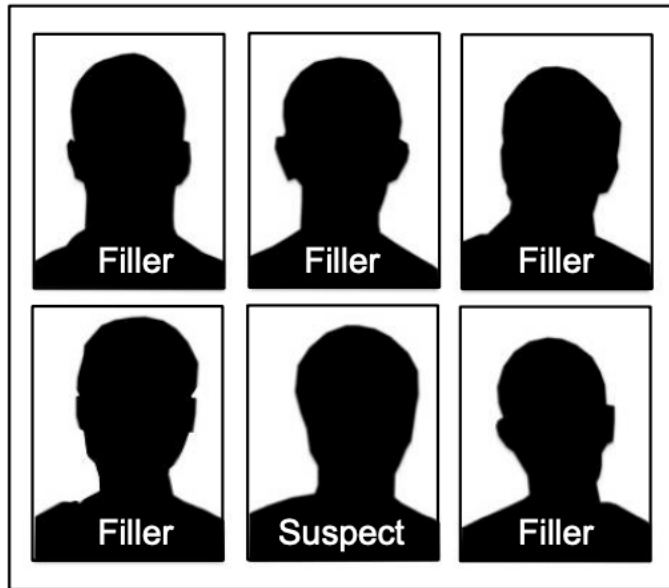


Figure 2 Variables that affect eyewitness accuracy categorized by stage of memory and by which stage of the crime/criminal proceedings each occur. Depending on the point in the proceedings, results from different analyses are needed to guide decisions. Note. *CAC* = confidence–accuracy characteristic; *ROC* = receiver operating characteristic.

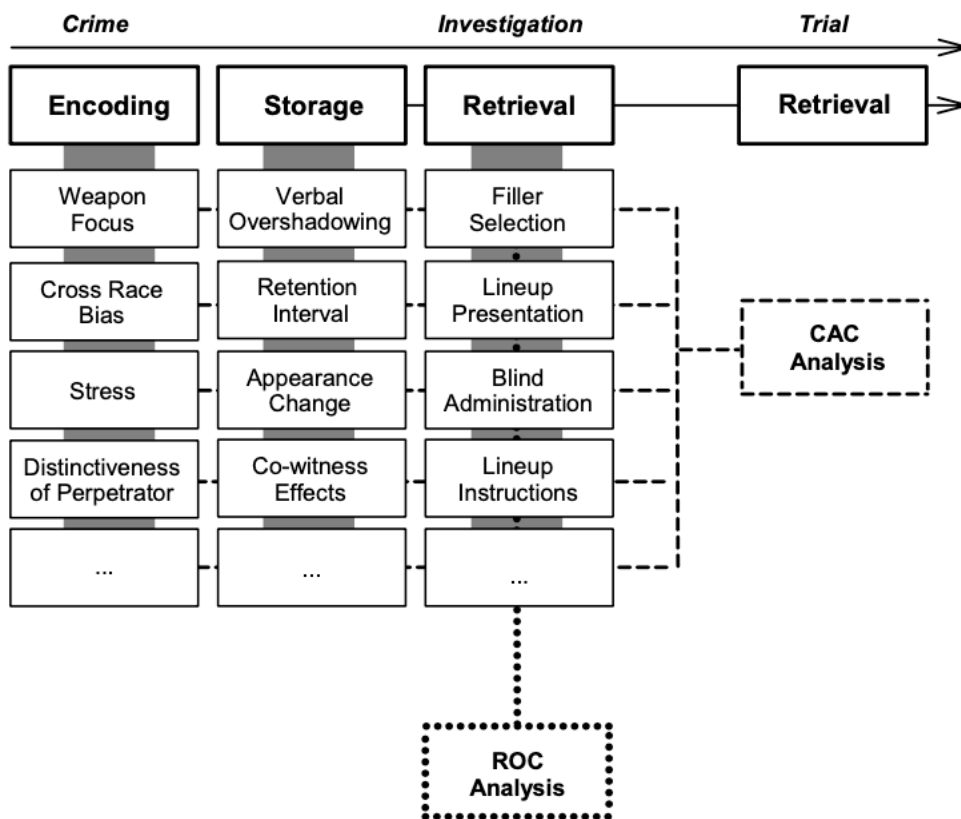


Figure 3 Hypothetical receiver operating characteristic data from two procedures; procedure A falls further from the line of chance performance (*dashed line*), therefore procedure A has better discriminability than procedure B. The *solid lines* represent the fit of a simple signal detection model (as shown in the top panel of Fig. 4). *pAUC*, partial area under the curve.

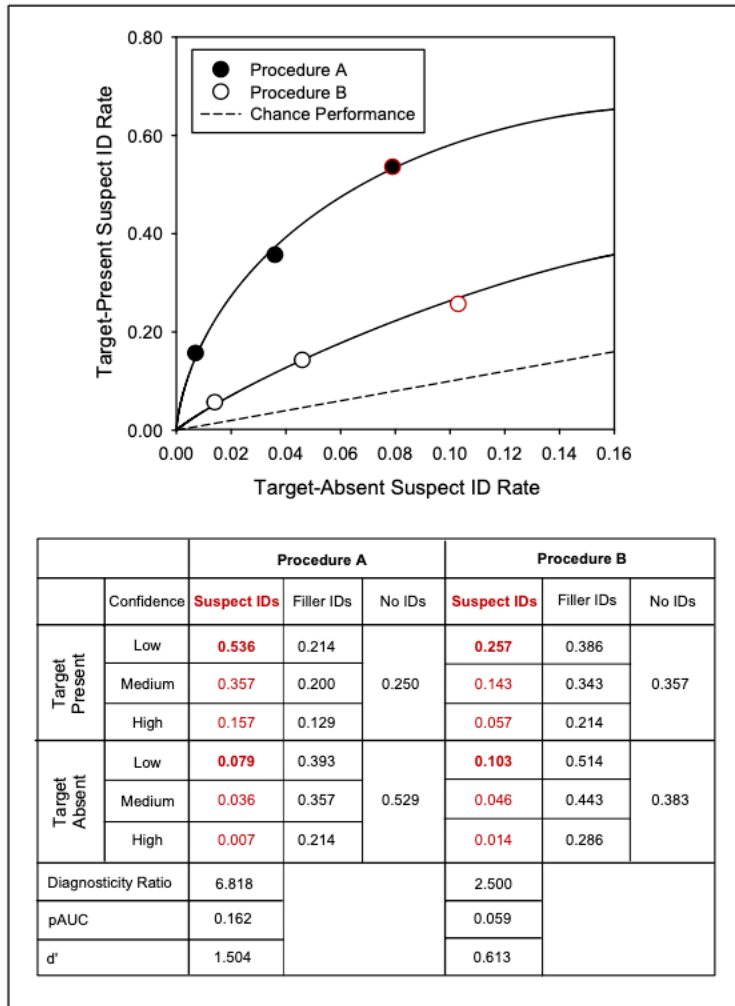
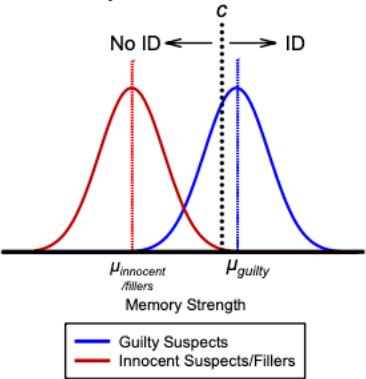


Figure 4 A simple signal detection-based model of a lineup task. The means (μ_{Innocent} , μ_{Fillers}) and standard deviations of the filler/innocent suspect distributions are 0 and 1, respectively. The mean (μ_{Guilty}) and standard deviation of the target distribution are d and 1, respectively. If the memory strength of the most familiar lineup member exceeds the criterion (c), an identification will be made. Unlike in a fair lineup, where fillers and innocent suspects are drawn from the same distribution (top panel on the left), in an unfair lineup, the fillers are drawn from a distribution that differs from the innocent suspect distribution (bottom panel on the left). In the unfair case, $\mu_{\text{Fillers}} < \mu_{\text{Innocent}}$ and results in a lower receiver operating characteristic (i.e., poorer ability to discriminate innocent from guilty suspects / poorer empirical discriminability) for unfair lineups (bottom right panel; see Colloff et al., 2016; 2018).

Fair Lineup



Unfair Lineup

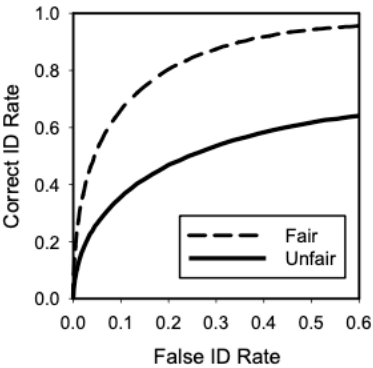
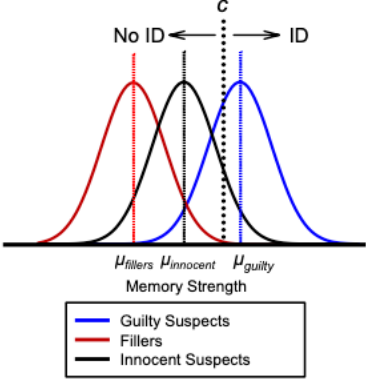


Figure 5 Two ways to measure the positive predictive value of an identification: calibration analysis (top figure), confidence–accuracy characteristic analysis (bottom figure),.

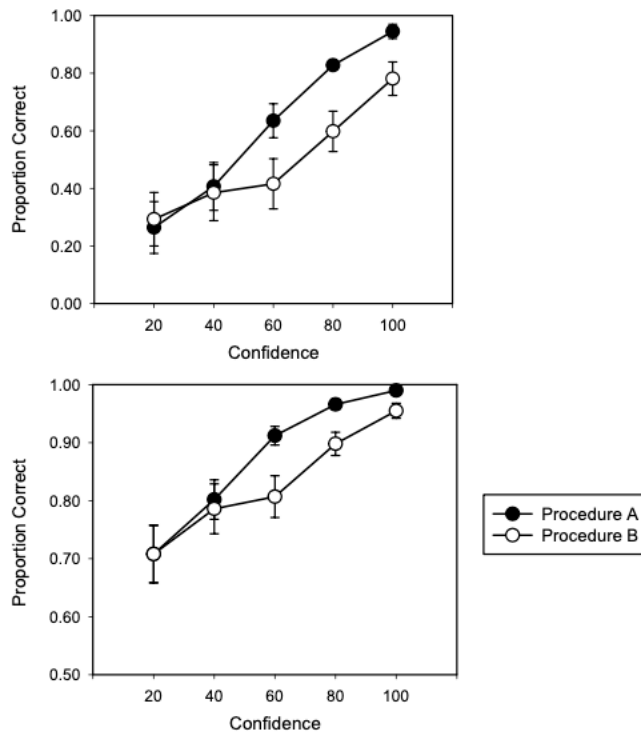


Table 1 Possible decisions and resulting outcomes

		Eyewitness Decision		
		Identified the Suspect	Identified a Filler	Did Not Identify Anyone
True State	Target Present	Correct ID (Hit)	Filler ID	Miss
	Target Absent	False ID (False Alarm)	Filler ID	Correct Rejection