

Review

The cognitive science of eyewitness memory

Laura Mickes^{1,*}, Brent M. Wilson², and John T. Wixted^{2,*}

Recent insights from cognitive science have reshaped our understanding of the reliability of eyewitness memory. Many believe that eyewitness memory is unreliable, but a better way of thinking is that eyewitness memory, like other types of forensic evidence, can be contaminated. Because contaminated evidence yields unreliable results, the focus should be placed on testing uncontaminated memory evidence collected early in a police investigation. The recent application of theories, principles, and methods from cognitive science has revealed that, both in the laboratory and in the real world, the first test of uncontaminated memory provides much more reliable information than was previously thought. Moreover, and crucially, this reliable but often-ignored evidence frequently points in the direction of a convicted defendant's innocence.

Highlights

Memory can be contaminated, exactly as other types of forensic evidence can be.

For that reason, eyewitness memory has long been thought to be unreliable.

However, the initial test of memory conducted early in a police investigation, which minimizes contamination, is more reliable than was previously believed.

On that all-important first test, eyewitnesses often provide reliable evidence that the police suspect is innocent.

Emerging insights into the reliability of eyewitness memory

Work by Elizabeth Loftus and her colleagues in the 1970s showed that eyewitness memory, like any type of forensic evidence, can be contaminated [1]. For example, a true memory of a stop sign can be replaced by a false memory of a yield sign based on nothing more than a passing suggestion [2]. Later, in the 1990s, it became clear that it is even possible for someone to acquire detailed and emotional false memories of traumatic events that never happened [3,4].

The susceptibility of human memory to contamination helps to explain a troubling pattern: hundreds of wrongful convictions, later overturned by DNA evidence, involved witnesses who confidently but incorrectly identified innocent defendants at trial¹. With no awareness of the insidious effects of memory contamination, these witnesses testified honestly, and the similarly unaware jurors naturally trusted that testimony to arrive at a guilty verdict. Because false memories look and feel as real as true memories [5,6], even skilled cross-examination by a defense attorney has little hope of shaking the faith of the witness in their memory of having seen the innocent defendant commit the crime. Revelations such as these have led many to question not only the value of eyewitness confidence but also the overall reliability of eyewitness memory.

In addition to the problem of memory contamination, a separate problem concerns the use of suggestive eyewitness identification procedures, which reveal to the witness who the police believe may have committed the crime [7]. Consider, for example, a **showup** (see [Glossary](#)) in which the police present a single suspect (sometimes in handcuffs) to a witness for a recognition decision. Under such conditions, it is obvious who the police suspect of being the perpetrator. Similar considerations apply to the ultimate suggestive identification test that later occurs at the criminal trial. The concern is that the witness may identify the suspect based on suggestive cues, not because the face of the suspect matches the witness's memory of the perpetrator.

To reduce suggestiveness and improve identification accuracy, several science-based guidelines have been proposed over the years, such as using a **pristine lineup** – usually in the form of a photo array [8,9]. In a typical 'six-pack' array, a photograph of the suspect is surrounded by photographs of five physically similar fillers, which allows the witness to make a recognition decision

¹School of Psychological Science, University of Bristol, Bristol, UK

²Department of Psychology, University of California, San Diego, CA, USA

*Correspondence:
laura.mickes@bristol.ac.uk (L. Mickes)
and jwixted@ucsd.edu (J.T. Wixted).



without knowing who is the suspect. To further improve identification accuracy, another recommendation focused on how the photographs are presented to the witness [10], either all at once (**simultaneous lineup**) or one at a time (**sequential lineup**). Studies conducted between 1985 and 2011 suggested that sequential lineups outperformed simultaneous lineups [11], and this body of work significantly influenced real-world police practices. For example, a survey of US law enforcement agencies conducted in 2013 found that approximately one-third had adopted the sequential procedure [12].

Although traditional eyewitness identification research has been influential, recent research grounded in cognitive science has reshaped our understanding of these issues. In what follows, we review some of the findings and analyses that led to the original conclusions and then review newer research indicating that (i) simultaneous lineups are diagnostically superior to sequential lineups (not the other way around), and (ii) confident identifications on an initial test of uncontaminated memory are more reliable than was previously believed. We then review evidence that witnesses who confidently misidentified an innocent defendant at trial, based on contaminated memory, often did otherwise on the initial test of their uncontaminated memory. Ironically, these new findings lead to an unexpected revelation: sometimes, eyewitness memory can help to reliably exonerate the wrongfully convicted.

Statistical measures supporting the traditional perspective

For a long time the field concluded that sequential lineups are diagnostically superior to simultaneous lineups based on a measure known as the **diagnosticity ratio**, which was the preferred measure of accuracy from the 1970s until about 2011. The basic finding was that sequential lineups substantially reduce false identifications of innocent suspects (the **false alarm rate**) while reducing correct identifications of guilty suspects (the **hit rate**) to a lesser degree. On balance, this is an intuitively appealing outcome. During this period, references to the 'sequential superiority effect' were common [11] and, in a survey of memory scientists published in 2001 [13], 63% of memory experts agreed that 'Witnesses are more likely to misidentify someone ... when presented with a simultaneous (as opposed to sequential) lineup.'

Over the same period of time, the confidence of an eyewitness was thought to be an unreliable indicator of accuracy. The field's generally pessimistic view of confidence was evident in the same 2001 survey [13], which found that >80% of experts endorsed the unqualified assertion that 'An eyewitness's confidence is not a good predictor of his or her identification accuracy.'

This negative assessment was based not only on the confident misidentifications known to have occurred at trial but also on the low correlation between eyewitness confidence and accuracy on an initial test of memory using a lineup. The correlation coefficient was the preferred confidence–accuracy measure for lineup data from the 1970s until about 1995 [14]. However, it is now well known that the correlation coefficient was the wrong measure to use because a low correlation can mask even a perfect confidence–accuracy relationship [15,16]. The field therefore switched to a more informative **calibration** approach which involves plotting accuracy separately for each level of confidence. Using this new approach, it became clear that the confidence of an eyewitness on an initial test of memory is, in truth, meaningfully related to accuracy [17].

Despite this new understanding, initial high-confidence identifications were still considered by key players to be highly error-prone. For example, in 2014, an authoritative amicus brief filed by the American Psychological Association stated the following: 'In one article reporting results from an empirical study, researchers found that among witnesses who made positive identifications, as many as 40 percent were mistaken, yet they declared themselves to be 90 percent to 100 percent

Glossary

Calibration: the alignment between the confidence of an individual in their judgments and the actual accuracy of those judgments. For example, judgments made with 80% confidence are well calibrated if they are accurate 80% of the time. In eyewitness identification research, correct judgments are identifications of guilty suspects, whereas incorrect judgments are identifications of innocent suspects and fillers.

Confidence–accuracy characteristic (CAC) analysis: a graph of the relationship between suspect ID accuracy and confidence. For each level of confidence, suspect ID accuracy is computed from the number of correct and incorrect suspect IDs made with that level of confidence. Specifically, suspect ID accuracy = correct suspect IDs divided by the number of correct plus incorrect suspect IDs.

Diagnosticity ratio: the hit rate divided by the false alarm rate.

Discriminability: the ability to distinguish between innocent and guilty suspects independently of response bias (often quantified using ROC analysis or d').

Discriminability index d' : a model-based measure of discriminability grounded in signal detection theory; it represents the degree to which underlying memory-match signals from two categories (e.g., innocent vs guilty suspects) overlap. For example, a d' of 1.5 indicates that the means of the two distributions of memory signals are 1.5 standard deviations apart.

Double-blind: both the officer administering the lineup and the witness are unaware of the identity of the suspect

False alarm rate: the proportion of TA lineups (or TA showups) presented to participants in which the innocent suspect is incorrectly identified.

Hit rate: the proportion of TP lineups (or TP showups) presented to participants in which the guilty suspect is correctly identified.

Miss rate: the proportion of TP lineups (or TP showups) presented to participants in which the guilty suspect is not identified.

Positive ID: eyewitness identification of the suspect from a showup or a lineup.

Posterior odds of guilt: the updated estimated odds that the suspect in a

confident in the accuracy of their identificationsⁱⁱ. The study highlighted in this amicus brief involved a low-stress simulated crime, followed by a pristine lineup procedure that tested the uncontaminated memory of each participant for a recently seen perpetrator [18]. Even under these ideal conditions, the calibration approach indicated that the accuracy of high-confidence identifications was only 60% correct.

To be sure, not everyone in the field shared this pessimistic perspective (e.g., [17,19]), but it was nevertheless the dominant perspective in expert surveys and consensus amicus briefs until at least 2014. Documents such as these inform courts of law about the field's understanding of the reliability of eyewitness memory. If eyewitness memory is that unreliable even on an initial test of uncontaminated memory conducted in accordance with best-practice recommendations under the low-stress conditions of a laboratory, it stands to reason that actual eyewitnesses tested under less than ideal conditions would be even less reliable.

Thus, circa 2012, the consensus was that sequential lineups are diagnostically superior to simultaneous lineups and that confident eyewitness identifications are unreliable even on an initial test of uncontaminated memory. However, the introduction of theories and methodologies from cognitive science would soon challenge this established perspective.

Reevaluating longstanding conclusions

A shift in thinking about eyewitness memory is reflected in a new survey of expert opinion [20]. This survey revealed that, although the opinions of scientists have remained largely unchanged on many issues (e.g., lineups should be **double-blind**), there has been a notable change in opinion about two issues: (i) the diagnostic accuracy of simultaneous versus sequential lineups, and (ii) the relationship between eyewitness confidence and accuracy. In contrast to the last survey in 2001, only 34% of surveyed experts now agree with the following statement: 'People are more accurate when making decisions about sequentially presented lineups compared with simultaneously presented lineups.' In addition, a remarkable 89% of experts now agree with the following statement about the confidence–accuracy relationship: "If best-practices are used during the lineup procedure, an eyewitness's confidence can provide information about the eyewitness's accuracy when obtained immediately after the identification decision."

Both these changes underscore how treacherous it can be to properly measure memory performance to answer seemingly simple questions such as 'Are sequential lineups diagnostically superior to simultaneous lineups?' and 'Is eyewitness confidence a reliable indicator of accuracy?' [21]. For example, in the absence of proper measurement methods grounded in a coherent model of underlying memory signals, it is deceptively easy to conflate **discriminability** and **response bias** (Box 1).

Simultaneous versus sequential lineups

As noted earlier, sequential lineups were long thought to be diagnostically superior to simultaneous lineups because, compared to simultaneous lineups, they reduce the false alarm rate (misidentifications of innocent suspects) more than they reduce the hit rate (correct identifications of guilty suspects). In other words, sequential lineups increase the diagnosticity ratio [11]. However, when hit and false alarm rates both change in the same direction, determining whether overall diagnostic accuracy has increased can be tricky. A more sophisticated methodology for addressing this issue is known as **receiver operating characteristic (ROC) analysis** [21–24].

When ROC analysis was first applied to this question, a different conclusion was reached [25]. Two other papers published in the same year applied signal detection-based approaches to

lineup is guilty after knowing the decision made by the eyewitness (e.g., the odds will increase if the witness identifies the suspect and will decrease if the witness instead rejects the lineup).

Prior odds of guilt: the estimated odds that the suspect in a lineup is guilty before knowing what the decision made by the eyewitness will be (e.g., the odds will be high if there is already considerable independent evidence that the suspect committed the crime).

Pristine lineup: a lineup, usually consisting of photographs, that contains only one suspect – the person the police believe may have committed the crime – and five or more known-to-be innocent fillers who, like the suspect, match the witness's description of the perpetrator (e.g., 'bald White male in his 30s with a goatee'). In addition, the police officer administering the lineup should be blind to the identity of the suspect to avoid unconsciously biasing the outcome, and the witness should be informed that the perpetrator may or may not be in the lineup.

Response bias: the inclination to make a positive identification (e.g., from a TA or TP lineup). As response bias changes from liberal to conservative, both the hit rate and the false alarm rate decrease.

Receiver operating characteristic (ROC) analysis: a plot of the hit rate versus the false alarm rate over various levels of response bias. The area under the curve provides a measure of how well participants can discriminate innocent from guilty suspects.

Sequential lineup: an identification procedure in which the lineup members (or their photographs or videos) are presented one at a time.

Showup: an identification procedure involving the suspect only (innocent or guilty).

Simultaneous lineup: an identification procedure in which the lineup members (or their photographs) are presented all at once.

Target-absent (TA) lineup: a lineup containing an innocent suspect along with several known-innocent fillers.

Target-present (TP) lineup: a lineup containing the guilty suspect along with several known-innocent fillers.

Box 1. Signal detection theory and ROC analysis

The main theoretical framework that has long been used to conceptualize recognition memory in the basic science literature is signal detection theory [86] (Figure I). As applied to eyewitness identification, the theory holds that the memory-match signals generated by guilty suspects in **target-present (TP) lineups** are relatively strong, on average, because the guilty suspect matches the eyewitness's memory of the person who committed the crime. The match will not always be perfect (e.g., the appearance of the guilty suspect may have recently changed), but, on average, the match should be stronger than that of an innocent face. Thus, across lineups, the distribution of memory signals generated by guilty suspects should have a higher mean than the distribution of memory signals generated by innocent suspects (and innocent fillers).

The ability of eyewitnesses to tell the difference between innocent and guilty suspects based on the strength of the underlying memory-match signal is called 'discriminability', and it is theoretically determined by the degree to which these two distributions overlap. Discriminability can be measured empirically using receiver operating characteristic (ROC) analysis (Figure II) or theoretically in terms of the underlying memory signals using the **discriminability index d'** .

Discriminability is distinct from response bias. Whether discriminability is low (e.g., for witnesses who only briefly observed the perpetrator) or high (e.g., for witnesses who had a long look at the perpetrator), response bias can range from liberal to conservative. For example, response bias would be liberal if the witnesses were instructed to make an ID from the lineup even if they have low confidence, and it would be conservative if the witnesses were instead instructed not to make an ID from the lineup unless they are 100% certain.

When liberal responding is encouraged, guilty suspects are more likely to be identified from TP lineups, thereby increasing the hit rate and decreasing the error of failing to identify a guilty suspect (**miss rate**). At the same time, however, innocent suspects are more likely to be misidentified from **target-absent (TA) lineups**, thus increasing the false alarm rate. There is thus a trade-off between the two errors. Because the optimal level of response bias is a subjective judgment call, the main goal of the relevant scientific research is to enhance discriminability. Viewing eyewitness identification research through the lens of signal detection theory is what changed the field's understanding of simultaneous versus sequential lineups and the confidence-accuracy relationship.

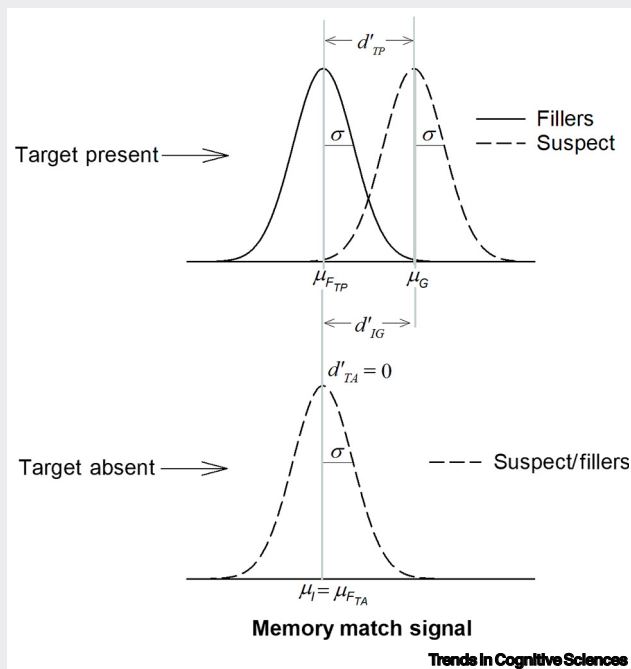


Figure I. Signal detection theory applied to police lineups. The simplest signal detection model of raw (untransformed) memory-match signals generated by faces in a fair lineup. A single face of the guilty suspect (TP) or innocent suspect (TA) is presented to the eyewitness along with innocent fillers for a decision. Across many lineups, the average memory-match signal generated by the guilty suspect (μ_G) is greater than the average memory-match signal generated by the innocent suspect (μ_I). Because the fillers are mnemonically equivalent to innocent suspects (i.e., they also match the witness's description of the perpetrator but do not otherwise match the memory of the perpetrator), the means of the distribution of memory signals they generate are the same as those of the innocent suspect (i.e., $\mu_{FTP} = \mu_{FTA} = \mu_I$). d'_{IG} is the standardized distance between the means of the innocent suspect and guilty suspect

distributions. Thus, $d'_{IG} = \frac{\mu_G - \mu_I}{\sigma}$, where σ represents the standard deviation of the relevant distributions. Two within-lineup discriminability measures are also relevant, d'_{TP} and d'_{TA} , that represent the standardized distance between the means of the suspect and filler distributions in TP and target-absent lineups, respectively. In a fair lineup, $d'_{TA} = 0$ because the innocent suspect is effectively only another filler. To make a decision, various confidence criteria are assumed to be ordered along the memory axis.

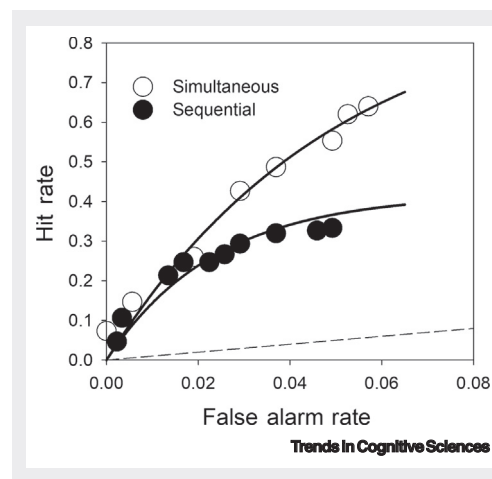


Figure II. Partial receiver operating characteristic (ROC) analysis. Empirical ROC data from a study comparing simultaneous and sequential lineups. Unlike standard ROCs, the axes do not range from 0 to 1.0, and this therefore is technically a partial ROC. The axes do not extend to 1.0 because even maximally liberal responding (i.e., every witness makes an identification from the lineup) would not result in the suspect being identified every time. This is true of both guilty suspects in TP lineups, and innocent suspects in TA lineups. For example, in fair TA lineups with six members, the identifications would land on the innocent suspect only 1/6 of the time, and the maximum false alarm rate is therefore $1/6 \approx 0.167$. Figure adapted, with permission, from experiment 1 in [25].

this topic and also challenged the idea that sequential presentation is superior [26,27]. Moreover, a subsequent review of the literature reported that, of 17 published and unpublished ROC comparisons, the simultaneous lineup was found to be diagnostically superior in 15, and significantly so in nine of them [28]. Measured using ROC analysis, a diagnostically superior procedure can achieve both a higher hit rate and a lower false alarm rate.

The ROC methodology used in these studies is grounded in a coherent model of underlying memory signals (Box 1), but some have argued against this approach and have advocated for a different type of ROC analysis grounded in applied considerations divorced from cognitive science [29,30]. We consider that approach to be unproductive [31], but the issue remains debated.

Why are simultaneous lineups diagnostically superior to sequential lineups? Multiple factors may play a role, including the 'stopping rule' that is often used in sequential lineups tested in the laboratory. That is, once a face is identified, the procedure stops (i.e., the remaining faces are not shown). This has the effect of empirically lowering the ROC [32,33]. Fortunately, police departments that have adopted the sequential procedure tend not to use a stopping rule.

However, there may be more to the story because a simultaneous advantage has sometimes been found even when no stopping rule is used (e.g., [34–36]). So far, only one cognitive theory at the level of underlying memory signals – known as diagnostic feature-detection theory [37,38] – has been advanced to explain why this might occur. According to this theory, simultaneous lineups immediately reveal to the witness that the faces in the lineup share many features (e.g., they are all White, male, about 30 years old, bald, and have a goatee). Shared features are nondiagnostic of innocence or guilt and therefore merely add noise to the decision-making process. Thus, excluding them from consideration should be diagnostically advantageous (*cf* [39]). Theoretically, when the faces are presented sequentially, it is not as readily apparent to the witness that the faces share nondiagnostic features, and the noise-adding features are more likely to be given weight, thus lowering the ROC (i.e., reducing discriminability).

Combining feature-matching logic with basic signal detection concepts has also led to new advances in other areas of eyewitness identification research. For example, a question that has bedeviled the field for decades concerns the optimal level of similarity between a suspect and the fillers in a lineup. They need to be somewhat similar, but if they are too similar, the task

becomes impossible [40]. What is the optimal level of similarity? Recent work grounded in cognitive science suggests an answer (Box 2).

Confidence and accuracy

The second issue that has seen a shift in thinking is the relationship between confidence and accuracy. It has long been known that signal detection theory predicts a strong confidence–accuracy relationship, and a strong relationship is typically observed in the basic science laboratory, such as when memory is tested using a list of words (e.g., [41–43]). Typically, recognition decisions made with high confidence are highly accurate, which initially raised a puzzle as to why high-confidence accuracy was so low when (i) the recognition task involved a lineup, and (ii) a calibration approach was used to analyze the data.

The puzzle was resolved when it was pointed out that the calibration approach inappropriately includes filler IDs (not only innocent suspect IDs) as errors when measuring accuracy for a given level of confidence [44]. Although filler IDs are errors, a better approach is to specifically measure suspect ID accuracy for each level of confidence because this measure addresses the question that is of interest to factfinders such as judges and jurors – given that the witness identified the suspect with a particular level of confidence, how accurate is that suspect ID likely to be? This approach is known as **confidence–accuracy characteristic (CAC) analysis**.

Another reason not to include filler IDs in the accuracy calculation is that suspect IDs and filler IDs have opposite implications for updating the **prior odds of guilt** [45,46]. Specifically, suspect IDs usually point in the direction of the guilt of the subject, whereas filler IDs usually point in the direction of the suspect's innocence. Thus, their implications for updating the prior odds of guilt should be considered separately, not combined. As shown in Figure 1, when the data are analyzed

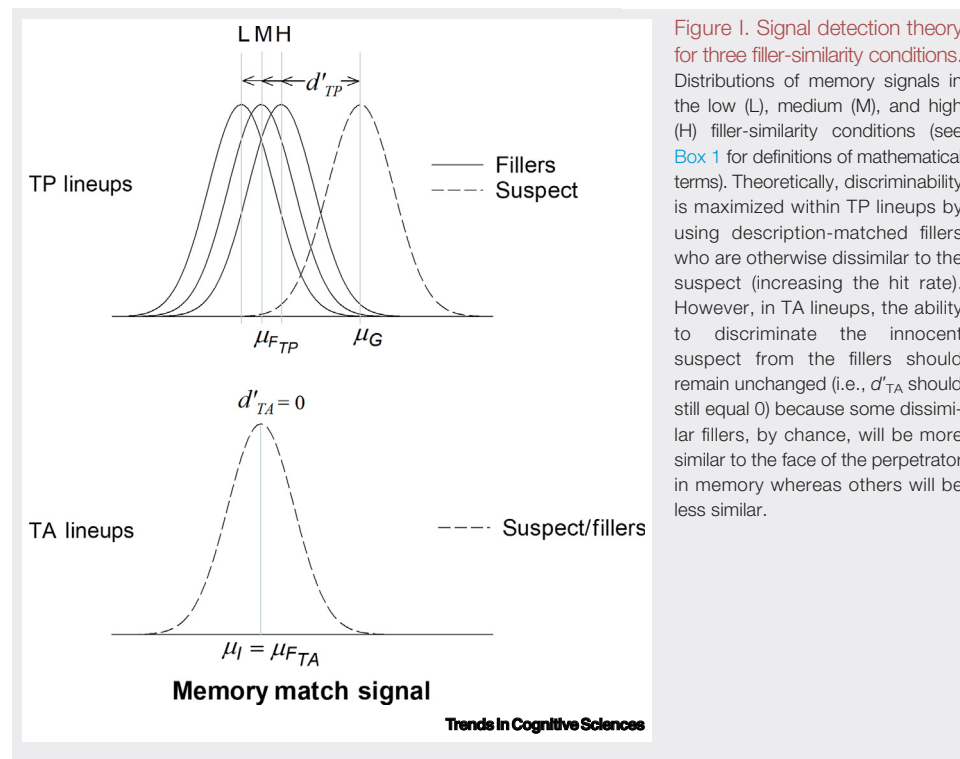
Box 2. Optimizing filler similarity in police lineups

How similar should the fillers in a lineup be to the suspect? [87]. A two-step procedure for optimizing the similarity of fillers to a suspect in a target-present (TP) or target-absent (TA) lineup was recently proposed [88–90]. The first step is to create a pool of description-matched fillers in accordance with current recommendations [9]. Thus, and crucially, any fillers drawn from this pool could be used to create a fair lineup. Nevertheless, within that pool, some fillers will be more similar to the suspect, and others will be less similar based on facial features that were not included in the witness's description of the perpetrator. Thus, filler similarity to the suspect can vary over a wide range even while retaining a fair, description-matched lineup.

In a TP lineup, the memory signals generated by high-similarity fillers would increase competition with the memory signal generated by the guilty suspect, thus reducing the hit rate. By contrast, the memory signals generated by low-similarity fillers would decrease competition with the memory signal generated by the guilty suspect, thereby increasing the hit rate. Thus, for TP lineups, maximizing filler dissimilarity would optimize performance by imperiling the guilty suspect.

In a TA lineup, by contrast, whether the fillers are high or low in similarity to the innocent suspect is largely orthogonal to their similarity to the perpetrator. Thus, the memory-match signals generated by high- and low-similarity fillers would be comparable. Because the use of low-similarity fillers in a TA lineup would not make the innocent suspect stand out, the false alarm rate (the probability of an innocent suspect being misidentified) would not be affected. As illustrated in Figure 1, it follows that the ability to discriminate innocent from guilty suspects would be maximized by the use of description-matched fillers (ensuring a fair lineup) who are otherwise maximally dissimilar to the suspect. The results were consistent with these predictions [88–90], confirming a long-overlooked result that had been reported decades earlier [40].

Another insight from cognitive science is that fillers should ideally be chosen not because they match the witness's description of the perpetrator but because they plausibly would have been described the same way [91]. A filler with a distinctive facial tattoo likely would not have been described as a 'bald White male in his 30s with a goatee', but he might nevertheless fit the description and would therefore be fair game using the standard match-to-description rule. However, such a filler would be properly excluded using this alternative rule.



separately in terms of suspect ID accuracy, confidence in a suspect ID is highly predictive of accuracy and, importantly, high-confidence suspect IDs are highly accurate [47,48].

It seems fair to say that these results constituted a surprising revelation about the reliability of high-confidence suspect IDs when memory is initially tested using a pristine lineup. In response, researchers have identified some moderators (e.g., a witness under marijuana intoxication) where the accuracy of high-confidence suspect identifications is lower than shown in Figure 1, even when using a pristine procedure [49–56]. This research is important because one or more

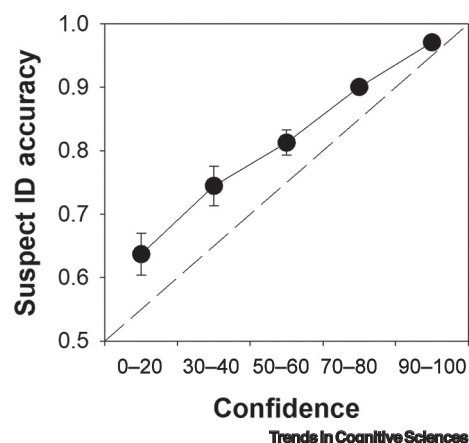


Figure 1. Eyewitness identification accuracy versus confidence. Confidence–accuracy characteristic (CAC) plot showing suspect ID accuracy (proportion correct) averaged across 15 studies with comparable scaling on the confidence (x) axis. Figure adapted, with permission, from [48].

moderators could be relevant to an actual criminal case. However, despite these exceptions, CAC analysis suggested that initial eyewitness identifications appear to be considerably more reliable than was once assumed.

Recall that the main reason for believing that eyewitnesses in the real world frequently make high-confidence errors on the initial test is because participants in the laboratory appeared to do so even under ideal conditions. In that case, one need not even bother checking to see whether eyewitnesses in the real world do the same on the initial test. However, on the initial test, eyewitnesses in the laboratory do not confidently misidentify suspects very often after all. It therefore makes sense to revisit the question of what actual eyewitnesses do on the first test.

Eyewitness identifications in the real world

In a police department field study of actual eyewitnesses tested using simultaneous or sequential lineups [57], the results indicated that high-confidence accuracy is impressively high, such that misidentifications of the innocent are relatively rare (Figure 2). In other words, these data from a field study largely correspond to laboratory data indicating that high-confidence IDs, especially when made quickly [58], are highly accurate. A more recent field study reinforced this perspective, and underscores the importance of both confidence and decision time [59].

The results shown in Figure 2 suggest that, even though some moderators may sometimes apply, high-confidence suspect IDs made from a pristine lineup tend to be accurate. This finding raises a question about whether eyewitnesses who confidently misidentify innocent defendants at trial also often do so on the initial test. It turns out that data bearing on this question had been presented earlier, but the implications were not immediately recognized.

An analysis of trial testimony in 161 DNA exoneration cases found that, in 57% of those cases (92 out of 161), what the eyewitness said about the defendant on the first test of their memory came up in trial testimony [60]. Remarkably, according to that testimony, in literally zero of those cases did the witness do on the first test what they would later do at trial, namely misidentify an innocent person immediately and with high confidence (the signature of a strong memory-match signal). This result accords with other evidence suggesting that when eyewitnesses are properly interviewed before memory contamination, their confident recollections are more reliable than was once believed [61,62].

To be sure, there are many DNA exoneration cases where we have no information about what happened on the first test, and the trial testimony we do have might not reflect what actually

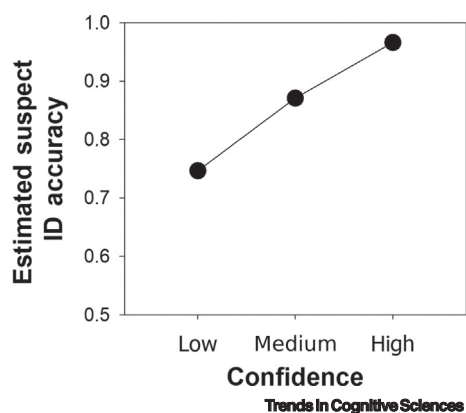


Figure 2. Confidence–accuracy relationship for a real-world field study. Estimated suspect ID accuracy (proportion correct) as a function of confidence for the data from the Houston Police Department field study assuming equal base rates. Suspect ID accuracy was estimated by fitting a signal detection model to the data. The same model was also used to estimate the base rate of guilty suspects in these police lineups. The estimate was 35% of lineups contained a guilty suspect and 65% contained an innocent suspect. This was the first model-based estimate of the base-rate of guilty suspects in actual police lineups. When this estimated base rate is assumed to be true, estimated accuracy remained high for high-confidence suspect IDs but was closer to 50% correct for low-confidence suspect IDs. Figure adapted, with permission, from [57].

happened. Nevertheless, our thinking should be guided by the information we have, not the information we do not have. The information we have about initial misidentifications in the real world converges with what was learned about participants in the laboratory: on the first test of uncontaminated memory, the evidence weighs against the idea that high-confidence misidentifications are common.

The remarkably consistent findings from laboratory studies, police department field studies, and DNA exoneration cases also square with what ought to be true from a theoretical perspective. From this perspective, an immediate ID made with high confidence occurs when the face of the suspect closely matches the face of the perpetrator represented in the brain of the eyewitness. Before memory contamination, an innocent suspect should rarely match the witness's memory of the perpetrator, which means that immediate high-confidence misidentifications of an innocent suspect should not be common at that time. This point applies even when a non-pristine, suggestive identification procedure is used, although such procedures do reduce high-confidence suspect ID accuracy to below the level displayed in [Figure 2 \[63,64\]](#).

A new perspective

These considerations call for a reevaluation of the reliability of eyewitness memory. The old message was that eyewitness memory is unreliable in the sense that high-confidence misidentifications are common on both the first test conducted early in a police investigation and the last test conducted at the criminal trial, irrespective of whether memory is contaminated or not. The new message is that eyewitness memory is much more reliable on the first test of uncontaminated memory than was previously believed.

Eyewitnesses often provide evidence of innocence on the first test

For a field that has long associated confident misidentifications with wrongful convictions, this is not an easy message to enthusiastically embrace. However, it accords with the relevant data and it leads to an unexpected revelation – one that can help to exonerate innocent prisoners. For many people currently serving long prison sentences, on the first (most reliable) test of memory, the same eyewitnesses whose confident trial testimony led to a guilty verdict provided reliable evidence of innocence. For example, on the first test, the witness either rejected the lineup (i.e., made no identification) or picked a filler. Contrary to intuition, these are not 'no information' outcomes. Instead, compared to whatever the prior odds of guilt might be, the **posterior odds of guilt** decrease.

The fact that a lineup rejection or a filler ID on the initial test is indicative of innocence has long been known [\[65–67\]](#), and recent work has investigated ways to further enhance evidence of innocence from a lineup ([Box 3](#)). However, in the decades since this point was first made the field has focused almost exclusively on the unreliability of eyewitness memory. The dominant message, by far, was 'Do not listen to what eyewitnesses have to say because they are too unreliable'. The new message is that eyewitness memory is not as unreliable as was once believed before memory contamination, and this is true whether the evidence points in the direction of innocence or guilt. Either way, it is important to listen to what the witness says on the first test – the one that minimizes the chances of testing contaminated memory. After all, with other types of forensic evidence, the focus is usually placed on evidence collected early in a police investigation precisely because it minimizes the chances that the evidence is contaminated. It should not be a radical proposition – or even a slightly controversial one – to suggest that the same principle applies to eyewitness memory.

In the DNA exoneration cases, the witnesses who made high-confidence errors at trial often got it right on the first test (before memory contamination). Crucially, this was true even though the police

Box 3. Enhancing evidence of innocence from police lineups

Figures 1 and 2 indicate that confidence can be a strong predictor of accuracy when a witness makes a positive ID from a lineup. Curiously, the opposite is often true when a lineup is rejected. That is, although a lineup rejection reduces the posterior odds of the suspect's guilt (providing evidence of innocence), confidence does not add a great deal of additional information. Why not? One notable difference between a positive ID and a lineup rejection is that confidence in a positive ID is made to a singular face, whereas confidence in a lineup rejection is made to the collective set of rejected faces. Simulations from a signal detection model indicated that confidence would more reliably indicate accuracy if, after a lineup rejection but before confidence is rated, the suspect is revealed to the witness, who is then asked – 'How sure are you that this person is not the perpetrator?' [92]. This approach was called the 'reveal procedure'. When confidence in a lineup rejection was applied only to the face of the subject, the confidence–accuracy relationship improved compared to a standard lineup procedure, and it did so in two ways. First, a high-confidence rejection became more indicative of innocence; second, many more innocent suspects received a high-confidence rejection. These findings were recently replicated independently [93].

The reveal procedure can be used whenever the lineup administrator is not blind to the identity of the suspect. In that case, the administrator is able to reveal the suspect to the witness. Related work investigated a similar procedure (the 'rule-out' procedure) that can be used when the lineup administrator is instead blind to the identity of the suspect. Using this procedure, if the suspect is not positively identified, the eyewitness is asked to rate their confidence for each lineup member they did not select (including the suspect), indicating how sure they are that each one is not the perpetrator [94,95]. As with the reveal procedure, this approach yields confidence that the individual suspect is not the perpetrator. The rule-out procedure enhances the confidence–accuracy relationship in the same way that the reveal procedure does.

The field has long been focused on the reliability of positive suspect IDs made from a lineup with high confidence. However, because many wrongful convictions began with a different initial identification outcome, research on those outcomes (including lineup rejections) should be a high priority.

in those cases rarely used pristine eyewitness identification procedures, as has been documented in detail [60]. To make this empirical observation is not to endorse the use of non-pristine identification procedures. After all, non-pristine procedures clearly do increase the likelihood of high-confidence misidentifications [48]. The point is that, even when a non-pristine procedure is used, eyewitnesses may not make an immediate high-confidence misidentification as often as was previously believed [68]. In fact, even then, eyewitnesses may often provide evidence of innocence (see [Outstanding questions](#)).

The initial identification test itself contaminates memory

Although not widely appreciated in the legal system, the first test of a witness's memory for a given suspect, whether pristine or not, changes (i.e., contaminates) the witness's memory for that suspect [9,69]. Indeed, decades of research in cognitive science have shown, conclusively, that testing memory by means of recognition ('Do you recognize this item as having been previously presented?') incidentally changes memory [70–73]. Surprisingly, making a recognition decision about a novel item creates a memory of it as effectively as intentionally encoding that item [70]. Note that the face of an innocent suspect in a lineup is a novel item, and this is why there is a new consensus recommendation to test a witness's memory of a suspect only once [9,69].

Naïve courtroom theories versus established mechanisms of memory

Unfortunately, courts of law are more comfortable dealing with the trial testimony of a witness who is under oath and available for cross-examination than they are with a one- or two-year-old police record of what the witness ostensibly said on the initial identification test. This might help to explain why lay theories of memory in the courtroom can easily take the focus off the first test, even when the inconsistent results of the first test are introduced at trial in the hope of impeaching the testimony of the eyewitness, as has happened in many of the DNA exoneration cases [60].

Naïve courtroom theories often endorse the idea that face recognition can fail on the first test but succeed with flying colors years later on a test at trial (e.g., 'I was nervous on the first test'). After

all, everyone has had the experience of failing to remember someone's name and then later remembering it accurately and with absolute certainty. However, thinking along these lines confuses memories that are retrieved by following a search process (recall of details) and memories characterized by the strength of a matching process (face recognition). Retrieving the name of a person involves a search process, and any search can fail on the first try and then succeed on a later try. By contrast, face recognition is a matching process that does not succeed or fail but instead yields a memory-match signal that ranges from weak to strong. Indeed, the major models of recognition memory in cognitive science are called global matching models [74–84].

A matching process occurs in a matter of seconds, but many wrongful convictions have happened even though the witness searched their memory for minutes on the first test before arriving at a low-confidence decision, as if face recognition is a search process. Such an outcome means that the faces in the lineup do not strongly match any face in the witness's memory, but if the witness happens to land on the suspect after minutes of deliberation, it might nonetheless be recorded as a **positive ID**. The face of the innocent suspect is now newly represented in the witness's brain, and it will therefore yield a stronger memory-match signal the next time memory for that same face is tested. This analysis from a cognitive science perspective also fits with a cognitive neuroscience perspective [85] (Box 4).

Concluding remarks

In a way, basic cognitive science is an applied science. Basic cognitive science creates a large repository of knowledge, some unpredictable subset of which will someday help to solve problems faced by society at large. The evolving understanding of eyewitness memory reviewed here provides one example because it has recently contributed to the exoneration of several innocent prisoners. We conclude by briefly telling the story of one of themⁱⁱⁱ.

In 1998, a young man named Miguel Solorio was suspected of having fired a shot during a confrontation between gangs that hit an elderly pedestrian, who later died from her wounds. The police placed his photograph in a lineup, but on the first test four witnesses rejected the lineup, providing reliable evidence of innocence. At his trial in 2000, two of those same witnesses confidently identified him, one with 100% certainty and the other with 75% certainty. Although there was no other evidence of guilt beyond the courtroom eyewitness identifications, Solorio was convicted and sentenced to life in prison without the possibility of parole. He was only 19 years old at the time of his arrest, and he spent 25 years of his life behind bars.

Based in part on the new understanding of the importance of focusing on the first test of a witness's uncontaminated memory, the Los Angeles County District Attorney's Office conceded that Solorio was innocent, and in November 2023 he was exonerated by a judge and released

Box 4. The cognitive neuroscience of face recognition

A basic tenet of cognitive neuroscience is that a memory record of a visual object is stored in the same regions of the brain that perceptually processed that object [96–98]. This means that, if the object in question happens to be a face, then the representation of that face is likely stored in the face-processing regions of the brain that initially supported face perception [97,99]. A key implication is that perceiving the face of an innocent suspect on an initial test will yield a weak memory-match signal (because the face has not been seen before) *and* will simultaneously create a memory record of that face, which may still be there the next time the same face is tested. By the time of the criminal trial, if the face of the innocent suspect has been seen multiple times (repeatedly refreshing the memory trace), it can generate a strong memory-match signal, leading to a confident misidentification. Moreover, no search of memory is required because the same features that initially drove neural activity in the face-processing regions of the brain (creating a memory record in those same neurons) will drive perception again. This time, however, the perceptual processing of the same face will come into contact with the contaminated memory record of the face.

Outstanding questions

How can lineup procedures be minimally modified (enhancing the prospects for real-world implementation) such that high-confidence rejections provide more compelling evidence of innocence? Confidence provides considerable information about the accuracy of a positive ID from a lineup, but the relationship is not as strong for lineup rejections (Box 3). Finding ways to strengthen it would help law enforcement to rule out innocent suspects early in a police investigation.

How can cognitive science be used to enhance evidence of guilt from a showup procedure? This is an important question to investigate because the police are often compelled to use a showup at an active crime scene. Compared to lineups, showups provide weaker evidence of guilt when the suspect is confidently identified.

How accurate are model-based estimates of the base rate of guilt from real-world police lineups, and what are the estimated base rates of guilt in different jurisdictions? Thus far, only one estimate has been obtained by fitting a signal detection model to data collected in the field.

Is there a better analytical method for comparing the diagnostic accuracy of competing lineup procedures than ROC analysis? Efforts have been made to improve upon lineup ROC analysis, but some have not been guided by a cognitive model of underlying memory signals. A main point of this paper is that progress is most likely when it is guided by principles and theories that are grounded in basic science.

How does the medial temporal lobe (MTL) interact with face-processing regions of the brain to encode and consolidate memories of a once-seen face? Research has found that faces are represented in the same regions of the brain that initially perceptually processed the faces. Separately, research has shown that MTL structures are important for storing and consolidating episodic memories. Investigating how these processes jointly operate would enhance our understanding of recognition memory for the once-seen face of a perpetrator.

from prison. Two other innocent prisoners who were exonerated a few months later (Abel Soto^{iv} and Jofama Coleman^v) were convicted for similar reasons. It seems almost certain that many more prisoners have suffered a similar fate, and there may be new hope for at least some of them.

Declaration of interests

The authors occasionally serve as paid expert witnesses in legal cases related to eyewitness memory.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT and Claude in order to improve language and readability. After using these AI tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Resources

ⁱ<https://innocenceproject.org/eyewitness-misidentification/>

ⁱⁱ<http://www.apa.org/about/offices/ogc/amicus/gomes-johnson>

ⁱⁱⁱ<https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=6701>

^{iv}<https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=6767>

^v<https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=6766>

References

- Loftus, E.F. and Palmer, J.C. (1974) Reconstruction of automobile destruction: an example of the interaction between language and memory. *J. Verbal Learn. Verbal Behav.* 13, 585–589
- Loftus, E.F. *et al.* (1978) Semantic integration of verbal information into a visual memory. *J. Exp. Psychol. Hum. Learn. Mem.* 4, 19–31
- Loftus, E.F. (1993) The reality of repressed memories. *Am. Psychol.* 48, 518–537
- Pendergrast, M. (2017) *The Repressed Memory Epidemic: How It Happened and What We Need To Learn From It*, Springer
- Bernstein, D.M. and Loftus, E.F. (2009) How to tell if a particular memory is true or false. *Perspect. Psychol. Sci.* 4, 370–374
- Wachendörfer, M.M. and Oeberst, A. (2023) Differences between true and false autobiographical memories: a scoping review. *Eur. Psychol.* 28, 247–264
- Wells, G.L. and Quinlivan, D.S. (2009) Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law Hum. Behav.* 33, 1–24
- Wells, G.L. *et al.* (1998) Eyewitness identification procedures: recommendations for lineups and photospreads. *Law Hum. Behav.* 22, 603–647
- Wells, G.L. *et al.* (2020) Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law Hum. Behav.* 44, 3–36
- Lindsay, R.C.L. and Wells, G.L. (1985) Improving eyewitness identifications from lineups: simultaneous versus sequential lineup presentation. *J. Appl. Psychol.* 70, 556–564
- Stebly, N.K. *et al.* (2011) Seventy-two tests of the sequential lineup superiority effect: a meta-analysis and policy discussion. *Psychol. Public Policy Law* 17, 99–139
- Police Executive Research Forum (2013) A National Survey of Eyewitness Identification Procedures in Law Enforcement Agencies. Published online June 2013. <https://www.ncjrs.gov/pdffiles1/nij/grants/242617.pdf>
- Kassin, S.M. *et al.* (2001) On the 'general acceptance' of eyewitness testimony research: a new survey of the experts. *Am. Psychol.* 56, 405–416
- Sporer, S.L. *et al.* (1995) Choosing, confidence, and accuracy: a meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychol. Bull.* 118, 315–327
- Juslin, P. *et al.* (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
- Roediger, H.L. *et al.* (2012) The curious complexity between confidence and accuracy in reports from memory. In *Memory and Law* (Nadel, L. and Sinnott-Armstrong, W., eds), pp. 84–118, Oxford University Press
- Brewer, N. and Wells, G.L. (2006) The confidence–accuracy relationship in eyewitness identification: effects of lineup instructions, foil similarity, and target-absent base rates. *J. Exp. Psychol. Appl.* 12, 11–30
- Sauer, J.D. *et al.* (2008) Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Legal Criminol. Psychol.* 13, 123–135
- Lindsay, D.S. *et al.* (1998) Accuracy and confidence in person identification: the relationship is strong when witnessing conditions vary widely. *Psychol. Sci.* 9, 215–218
- Seale-Carlisle, T.M. *et al.* (2024) New insights on expert opinion about eyewitness memory research. *Perspect. Psychol. Sci.*, Published online April 18, 2024. <https://doi.org/10.1177/17456916241234837>
- Brady, T.F. *et al.* (2023) Measuring memory is harder than you think: how to avoid problematic measurement practices in memory research. *Psychon. Bull. Rev.* 30, 421–449
- Gronlund, S.D. *et al.* (2012) Showups versus lineups: an evaluation using ROC analysis. *J. Appl. Res. Mem. Cogn.* 1, 221–228
- Gronlund, S.D. *et al.* (2014) Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Curr. Dir. Psychol. Sci.* 23, 3–10
- Wixted, J.T. and Mickes, L. (2018) Theoretical vs. empirical discriminability: the application of ROC methods to eyewitness identification. *Cogn. Res. Princ. Implic.* 3, 9
- Mickes, L. *et al.* (2012) Receiver operating characteristic analysis of eyewitness memory: comparing the diagnostic accuracy of simultaneous and sequential lineups. *J. Exp. Psychol. Appl.* 18, 361–376
- Clark, S.E. (2012) Costs and benefits of eyewitness identification reform: psychological science and public policy. *Perspect. Psychol. Sci.* 7, 238–259
- Palmer, M.A. and Brewer, N. (2012) Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law Hum. Behav.* 36, 247–255
- Mickes, L. and Wixted, J.T. (2024) Eyewitness memory. In *Oxford Handbook of Human Memory* (Kahana, M.J. and Wagner, A.D., eds), pp. 2294–2328, Oxford University Press
- Smith, A.M. *et al.* (2020) Distinguishing between investigator discriminability and eyewitness discriminability: a method for

- creating full receiver operating characteristic curves of lineup identification performance. *Perspect. Psychol. Sci.* 15, 589–607
30. Smith, A.M. *et al.* (2022) Fair lineups improve outside observers' discriminability, not eyewitnesses' discriminability: evidence for differential filler-siphoning using empirical data and the WITNESS computer-simulation architecture. *J. Appl. Res. Mem. Cogn.* 11, 534
 31. Wilson, B.M. *et al.* (2024) Feeling the future of eyewitness research. *Cognition* 251, 105879
 32. Rotello, C.M. and Chen, T. (2016) ROC curve analyses of eyewitness identification decisions: an analysis of the recent debate. *Cogn. Res. Princ. Implic.* 1, 10
 33. Wilson, B.M. *et al.* (2019) Making sense of sequential lineups: an experimental and theoretical analysis of position effects. *J. Mem. Lang.* 104, 108–125
 34. Seale-Carlisle, T.M. and Mickes, L. (2016) US line-ups outperform UK line-ups. *R. Soc. Open Sci.* 3, 160300
 35. Seale-Carlisle, T.M. *et al.* (2019) Designing police lineups to maximize memory performance. *J. Exp. Psychol. Appl.* 25, 410–430
 36. Ayala, N.T. *et al.* Beyond the confidence–accuracy relation: a multiple-reflector-variable approach to postdicting accuracy on eyewitness lineups. *J. Exp. Psychol. Appl.* (in press)
 37. Wixted, J.T. and Mickes, L. (2014) A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychol. Rev.* 121, 262
 38. Carlson, C.A. *et al.* (2019) Lineup fairness: propitious heterogeneity and the diagnostic feature-detection hypothesis. *Cogn. Res. Princ. Implic.* 4, 20
 39. Tversky, A. (1977) Features of similarity. *Psychol. Rev.* 84, 327–352
 40. Wells, G.L. *et al.* (1993) The selection of distractors for eyewitness lineups. *J. Appl. Psychol.* 78, 835
 41. Mickes, L. *et al.* (2011) Strong memories are hard to scale. *J. Exp. Psychol. Gen.* 140, 239–257
 42. Tekin, E. *et al.* (2021) Applying confidence accuracy characteristic plots to old/new recognition memory experiments. *Memory* 29, 427–443
 43. Tekin, E. and Roediger, H.L. (2017) The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cogn. Res. Princ. Implic.* 2, 49
 44. 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
 45. Wixted, J.T. (2022) The basic science of eyewitness identification. In *Visual Memory* (Brady, T.F. and Bainbridge, W.A., eds), pp. 315–334, Routledge
 46. Wixted, J.T. and Roediger, H.L. (2023) Signal detection theory and eyewitness identification. In *Memory in Science for Society: There Is Nothing as Practical as a Good Theory* (Logie, R.H. *et al.*, eds), pp. 63–91, Oxford University Press
 47. Wixted, J.T. *et al.* (2015) Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *Am. Psychol.* 70, 515–526
 48. Wixted, J.T. and Wells, G.L. (2017) The relationship between eyewitness confidence and identification accuracy: a new synthesis. *Psychol. Sci. Public Interest* 18, 10–65
 49. Carlson, M.A. *et al.* (2023) The sleepy eyewitness: self-reported sleep predicts eyewitness memory. *J. Appl. Res. Mem. Cogn.* 12, 513–530
 50. Carlson, C.A. *et al.* (2023) Was he the perpetrator or a bystander? Testing theories of unconscious transference for eyewitness identification. *J. Appl. Res. Mem. Cogn.* 13, 543–555
 51. Giacona, A.M. *et al.* (2021) Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law Hum. Behav.* 45, 256–270
 52. Grabman, J.H. *et al.* (2019) Predicting high confidence errors in eyewitness memory: the role of face recognition ability, decision-time, and justifications. *J. Appl. Res. Mem. Cogn.* 8, 233–243
 53. Lockamy, R.F. *et al.* (2020) The effect of viewing distance on empirical discriminability and the confidence–accuracy relationship for eyewitness identification. *Appl. Cogn. Psychol.* 34, 1047–1060
 54. Moore, K.N. *et al.* (2024) Eyewitness confidence does not necessarily indicate identification accuracy. *Policy Insights Behav. Brain Sci.* 11, 149–155
 55. Pezdek, K. *et al.* (2020) Marijuana impairs the accuracy of eyewitness memory and the confidence–accuracy relationship too. *J. Appl. Res. Mem. Cogn.* 9, 60–67
 56. Sauer, J.D. *et al.* (2019) Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychol. Public Policy Law* 25, 147–165
 57. Wixted, J.T. *et al.* (2016) Estimating the reliability of eyewitness identifications from police lineups. *Proc. Natl. Acad. Sci. U. S. A.* 113, 304–309
 58. Seale-Carlisle, T.M. *et al.* (2019) Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world. *J. Appl. Res. Mem. Cogn.* 8, 420–428
 59. Quigley-McBride, A. and Wells, G.L. (2023) Eyewitness confidence and decision time reflect identification accuracy in actual police lineups. *Law Hum. Behav.* 47, 333–347
 60. Garrett, B. (2011) *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*, Harvard University Press
 61. Spearing, E.R. and Wade, K.A. (2022) Providing eyewitness confidence judgments during versus after eyewitness interviews does not affect the confidence–accuracy relationship. *J. Appl. Res. Mem. Cogn.* 11, 54–65
 62. Wixted, J.T. *et al.* (2018) Rethinking the reliability of eyewitness memory. *Perspect. Psychol. Sci.* 13, 324–335
 63. Eisen, M.L. *et al.* (2017) An examination of showups conducted by law enforcement using a field-simulation paradigm. *Psychol. Public Policy Law* 23, 1–22
 64. Eisen, M.L. *et al.* (2022) Comparing witness performance in the field versus the lab: how real-world conditions affect eyewitness decision-making. *Law Hum. Behav.* 46, 175–188
 65. Wells, G.L. and Lindsay, R.C. (1980) On estimating the diagnosticity of eyewitness nonidentifications. *Psychol. Bull.* 88, 776–784
 66. Wells, G.L. *et al.* (2015) Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law Hum. Behav.* 39, 99–122
 67. Clark, S.E. *et al.* (2008) Regularities in eyewitness identification. *Law Hum. Behav.* 32, 187–218
 68. Mickes, L. *et al.* (2017) Distilling the confidence–accuracy message: a comment on Wixted and Wells (2017). *Psychol. Sci. Public Interest* 18, 51–56
 69. Wixted, J.T. *et al.* (2021) Test a witness's memory for a suspect only once. *Psychol. Sci. Public Interest* 22, 1S–18S
 70. Buckner, R.L. *et al.* (2001) Encoding processes during retrieval tasks. *J. Cogn. Neurosci.* 13, 406–415
 71. Criss, A.H. *et al.* (2018) Release from output interference in recognition memory: a test of the attention hypothesis. *Q. J. Exp. Psychol.* 71, 1081–1089
 72. Jacoby, L.L. *et al.* (2005) Modes of cognitive control in recognition and source memory: depth of retrieval. *Psychon. Bull. Rev.* 12, 852–857
 73. Velanova, K. *et al.* (2003) Functional-anatomic correlates of sustained and transient processing components engaged during controlled retrieval. *J. Neurosci.* 23, 8460–8470
 74. Cox, G.E. and Shiffrin, R.M. (2017) A dynamic approach to recognition memory. *Psychol. Rev.* 124, 795–860
 75. Clark, S.E. (2003) A memory and decision model for eyewitness identification. *Appl. Cogn. Psychol.* 17, 629–654
 76. Dennis, S. and Humphreys, M.S. (2001) A context noise model of episodic word recognition. *Psychol. Rev.* 108, 452–478
 77. Gillund, G. and Shiffrin, R.M. (1984) A retrieval model for both recognition and recall. *Psychol. Rev.* 91, 1–67
 78. Hintzman, D.L. (1986) 'Schema abstraction' in a multiple-trace memory model. *Psychol. Rev.* 93, 411–428
 79. Hintzman, D.L. (1988) Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychol. Rev.* 95, 528–551
 80. McClelland, J.L. and Chappell, M. (1998) Familiarity breeds differentiation: a subjective-likelihood approach to the effects of experience in recognition memory. *Psychol. Rev.* 105, 724–760
 81. Murdock, B.B. (1982) A theory for the storage and retrieval of item and associative information. *Psychol. Rev.* 89, 609–626
 82. Murdock, B.B. (1993) TODAM2: a model for the storage and retrieval of item, associative, and serial-order information. *Psychol. Rev.* 100, 183–203

83. Osth, A.F. and Dennis, S. (2015) Sources of interference in item and associative recognition memory. *Psychol. Rev.* 122, 260–311
84. Shiffrin, R.M. and Steyvers, M. (1997) A model for recognition memory: REM – retrieving effectively from memory. *Psychon. Bull. Rev.* 4, 145–166
85. Wixted, J.T. (2024) The cognitive neuroscience of eyewitness identification. In *Forensic Psychology: Crime, Justice, Law, Interventions* (Davies, G.M. et al., eds), pp. 723–726, Wiley
86. Green, D.M. and Swets, J.A. (1966) *Signal Detection Theory and Psychophysics*, John Wiley
87. Fitzgerald, R.J. et al. (2015) Suspect filler similarity in eyewitness lineups: a literature review and a novel methodology. *Law Hum. Behav.* 39, 62–74
88. Colloff, M.F. et al. (2021) Optimizing the selection of fillers in police lineups. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2017292118
89. Shen, K.J. et al. (2023) Modeling face similarity in police lineups. *Psychol. Rev.* 130, 432
90. Shen, K.J. et al. (2024) The effects of filler similarity and lineup size on eyewitness identification. *J. Exp. Psychol. Learn. Mem. Cogn.* 50, 1444–1462
91. Frank, M.C. and Goodman, N.D. (2012) Predicting pragmatic reasoning in language games. *Science* 336, 998
92. Yilmaz, A.S. et al. (2022) The reveal procedure: a way to enhance evidence of innocence from police lineups. *Law Hum. Behav.* 46, 164–173
93. Seale-Carlisle, T.M. (2024) Improving the diagnostic value of lineup rejections. *Cognition* 252, 105917
94. Ayala, N.T. et al. (2022) The rule-out procedure: increasing the potential for police investigators to detect suspect innocence from eyewitness lineup procedures. *J. Appl. Res. Mem. Cogn.* 11, 489
95. Smith, A.M. et al. (2023) The rule-out procedure: a signal-detection-informed approach to the collection of eyewitness identification evidence. *Psychol. Public Policy Law* 29, 19
96. Brodt, S. and Gais, S. (2021) Memory engrams in the neocortex. *Neuroscientist* 27, 427–444
97. She, L. et al. (2024) Temporal multiplexing of perception and memory codes in IT cortex. *Nature* 629, 861–868
98. Slotnick, S.D. and Schacter, D.L. (2004) A sensory signature that distinguishes true from false memories. *Nat. Neurosci.* 7, 664–672
99. Kanwisher, N. and Yovel, G. (2006) The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 361, 2109–2128