Chapter 4: The Case of Confidence in Eyewitness Testimony

Kyros J. Shen, Allan L. Lam & John T. Wixted

University of California, San Diego

John T. Wixted, Department of Psychology, University of California, San Diego, jwixted@ucsd.edu.

Abstract

In this chapter, we discuss the complicated history of the field's understanding of how confidence in an eyewitness identification of a suspect is related to accuracy. In part 1, we discuss how the misconception that eyewitness identifications are inherently unreliable–regardless of confidence–came to be. It is a story that has several interrelated themes: (1) DNA exoneration cases that frequently involved eyewitness misidentifications, (2) lab-based research showing how malleable memory is, and (3) lab studies concluding that eyewitness confidence is, at best, only weakly related to accuracy. In part 2, we discuss how that verdict has been overturned. As it turns out, just as with other kinds of forensic evidence, there are conditions under which eyewitness memory is reliable and conditions under which it is unreliable. Recent evidence indicates that an initial, immediate, high-confidence eyewitness identification from a proper lineup is highly reliable in both laboratory and real-world settings. In part 3, we review basic science showing why it is essential to keep the focus on the initial test of a witness's memory because it not only minimizes the chances of testing contaminated memory but the test itself also contaminates the witness's memory for the suspect. Finally, in part 4, we consider a real-world case illustrating how the legal system goes astray by focusing on the last test of memory, at trial (thereby maximizing the chances of testing contaminated memory) instead of focusing on the all-important first test (thereby minimizing the chances of testing contaminated memory).

The Case of Confidence in Eyewitness Testimony

To assess the reliability of eyewitness testimony, it is essential to consider the relationship between the subjective assessment of an eyewitness's likelihood of making a correct identification (*confidence*) and objective performance (*accuracy*). However, when asked how reliable eyewitness memory is, most people reply with answers like “not very reliable” or “not reliable at all.” This understandable response accords with the large number of wrongful convictions associated with confident eyewitness misidentifications in a court of law. According to the Innocence Project, eyewitness misidentifications played a role in almost 70% of 375 wrongful convictions overturned by DNA evidence since 1989, which is more than any other contributing factor (Innocence Project, 2023).

In addition to these cold statistics, highly publicized real-life cases have further underscored the unreliability of eyewitness memory, even when an eyewitness exhibited the highest level of certainty at the criminal trial. The case of Ronald Cotton is one of the most well-known examples, and his story was eventually published as a compelling memoir called *Picking Cotton* (Cotton, Thompson-Cannino, & Torneo, 2009). In 1984, Jennifer Thompson, a college student at the time, was brutally raped, and Cotton was wrongfully convicted based mainly on her eyewitness identification. Thompson was absolutely certain that she made the correct identification, but, after serving more than 10 years in prison, Cotton was eventually exonerated based on DNA evidence that revealed the identity of the true rapist (a man named Bobby Poole).

Although real-life examples like this helped to create the widespread impression that eyewitness memory is unreliable, it was scientific research conducted by experimental psychologists that fully cemented that impression in the minds of many. This research was regarded as being sufficiently conclusive and sufficiently accepted by the field at large that it was described in amicus briefs filed by the American Psychological Association as follows: “The findings drawn from this research are also overwhelmingly accepted within the scientific community” (American Psychological Association, 2014a, p. 4). In this same brief, one of the most widely accepted findings was summarized as follows: “This belief that confidence correlates with accuracy was once shared by some courts. See, e.g., Neil v. Biggers, 409 U.S. 188, 199 (1972). Modern scientific research, however, has called that notion into very serious question” (American Psychological Association, 2014a, p. 6).

If eyewitness memory is error prone, and if the relationship between confidence and accuracy is weak, then the implication would be that eyewitness memory is inherently unreliable. Indeed, in a subsequent APA amicus brief filed in 2016, the upshot of the overall message from the psychological science of eyewitness identification was characterized this way: “Moreover, although the unreliability of eyewitness identifications is well known in the scientific community and among many lawyers, it is not understood by lay juries (American Psychological Association, 2016, p. 9).” And as recently as 2019, an amicus brief doubled down on this pessimistic message, emphasizing the fact that “eyewitnesses are frequently mistaken in their identifications” (American Psychological Association, 2019, p. 6).

Fast forward to today, and we can now find statements such as “our understanding of the relation between confidence and accuracy in eyewitness identification has changed… psychological studies increasingly favor the idea that confidence is closely related to accuracy under some conditions and poorly related under others” (Schacter, 2022, p. 452) and “Under those conditions, … (a) confidence and accuracy are strongly related and (b) high-confidence suspect identifications are remarkably accurate” (Wixted & Wells, 2017, p. 10). In other words, researchers now hold a much more nuanced opinion when discussing the confidence-accuracy relationship in eyewitness memory.

What caused opinion to shift away from the almost universal belief about how unreliable eyewitness memory is to the more nuanced understanding that it is sometimes highly reliable? This chapter will discuss the field's complicated journey to this new understanding and how, counterintuitively, it provides actual evidence of innocence for many prisoners who may have been wrongfully convicted but for whom no DNA evidence will ever be available to exonerate them.

Part 1: The History of Eyewitness Memory Research

It is important to emphasize that the DNA exonerations occurred in the context of scientific research that was already casting serious doubt on the reliability of eyewitness memory. For example, the 1970s and 1980s saw an upsurge in false memory studies calling into question the intrinsic fidelity of human memory. Loftus and Palmer's (1974) groundbreaking work set the stage by demonstrating the potential malleability of memories as a result of post-event misleading information. Their seminal experiment, which is covered in virtually every psychology textbook, demonstrated how merely tweaking a suggestive word ("smashed" versus "contacted") in a post-event questionnaire could alter eyewitnesses' recollections, leading them to estimate the cars traveling at higher speeds and even falsely remember the existence of broken glass on the ground.

Loftus et al. (1978) further highlighted the susceptibility of visual memories to post-event suggestions by leading participants to mistakenly remember a stop sign instead of the original yield sign from a traffic accident scene. Although there are controversies surrounding this memory phenomenon (McCloskey & Zaragoza, 1985; Zaragoza et al., 1987), the result itself was consistent across a number of different studies. This consistency persisted even as studies ventured into the realm of much more complex fabricated memories, such as the "lost in a shopping mall" scenario (Loftus & Pickrell, 1995), where a false memory of getting lost in the mall as a child was implanted in nearly 30% of the participants tested. Another example is the "hot air balloon ride" (Wade et al., 2002), where the false memory of riding a hot air balloon as a child was induced with a fake photograph. Even emotional false memories involving crimes have been successfully induced (Shaw & Porter, 2015).

The inescapable message from the false memory research conducted by Loftus and her colleagues was that, far more than had been previously appreciated, memory is *malleable*. In other words, just as other kinds of forensic evidence can be contaminated (e.g., DNA, fingerprints), so too can an eyewitness's recall of events. A separate line of research suggested that the same might be true of eyewitness identification, which is a recognition test. A recognition test does not require the recall of details but instead asks whether a physically present stimulus (such as a face in a lineup) was seen before. For example, Luus & Wells (1994) found that eyewitnesses to a staged crime became significantly more confident in their identifications when told that a co-eyewitness picked the same suspect, and significantly less confident when told that the co-eyewitness either did not make an ID or picked a different suspect. Wells & Bradfield (1998) later demonstrated that eyewitnesses who received confirming feedback after making the initial identification subsequently reported inflated levels of confidence in their decision.

These revelations about the malleability of memory (and corresponding malleability of confidence) raised two major concerns. First, juries find confident eyewitness identifications to be extremely compelling. Second, there is usually a long gap between the crime and the trial, a period where eyewitness memories are continuously vulnerable to distortion. Thus, for decades, and still to this day, eyewitness memory researchers have cautioned the legal system about placing much weight on the confidence a witness expresses at trial when identifying a defendant in front of a jury. This message has not changed over time.

If eyewitness memory is unreliable after it has been contaminated (just every other form of forensic evidence is when contaminated), it seems natural to wonder about how reliable it is—and whether confidence is related to accuracy—before it is contaminated. In other words, in a test of uncontaminated memory using a proper lineup procedure (e.g., a fair lineup, see next section), is eyewitness confidence in a positive identification strongly related to the accuracy of that identification? If eyewitness memory is unreliable even under those conditions, then, obviously, relying on eyewitness memory in forensically relevant settings would never make sense. In that case, the problem would be that eyewitness memory is simply an inherently unreliable form of evidence, and the job of scientists would be to make this clear to the legal system.

In the early research investigating this issue, scientists took an intuitive approach by computing the correlation between eyewitness confidence and accuracy. After all, if you want to find out if two measures are related (in this case, confidence and accuracy), then computing the correlation between them seems like a reasonable thing to do. When a lineup is used to test memory, the specific correlational measure of interest is called the point-biserial correlation coefficient (r_{pb}). This measure represents the correlation between the dichotomous outcome of accuracy (each participant is either correct or incorrect in their lineup decision) and the continuous measure of self-reported confidence (i.e., each participant also supplies a confidence rating). This approach attempts to boil the confidence-accuracy relationship down to a single number.

Early studies using this method often reported low r_{pb} values of ~.20, (Deffenbacher, 1980; Bothwell, Deffenbacher & Brigham, 1987; Wells & Murray 1984), though somewhat higher values closer to ~.40 were obtained when the analysis was limited to “choosers,” that is, to participants who did not reject the lineup (Sporer et al., 1995). These low point-biserial correlations, which were obtained under ideal testing conditions in the lab, understandably created considerable skepticism toward eyewitness confidence in general. If the correlation is that low even under perfect testing conditions in the lab and even when memory has not yet been contaminated, it seems reasonable to suppose that the correlation is essentially negligible in the real world.

Putting it all together yields a rather damning picture of the confidence-accuracy relationship and of the reliability of eyewitness memory in general. First, the DNA exoneration statistics revealed the outsized role played by eyewitness misidentifications in wrongful convictions. Second, influential scientific research showed how malleable human memory can be. And third, another line of scientific research appeared to show how little an eyewitness’s expressed confidence tells us about their accuracy, even under the best of conditions. All of this led to a clear verdict: eyewitness memory is simply unreliable, and the confidence expressed by an eyewitness, although convincing to juries, should be disregarded.

It is perhaps not surprising, then, that in a survey that summarized the opinions of 63 eyewitness experts, more than 80% agreed that confidence is not a good predictor of accuracy (Kassin, Ellsworth, & Smith, 1989). In a later survey, Kassin et al. (2001) again found that more than 80% of experts still agreed with the statement that “An eyewitness’s confidence is not a good predictor of his or her identification accuracy.” A remarkable 87% of experts agreed with that statement, which is even more than the 83% who agreed with a statement summarizing one of the most well-established principles in all of experimental psychology: “The rate of memory loss for an event is greatest right after the event and then levels off over time.” The fact that typical time course of forgetting follows that basic trajectory, now enshrined in the “power law of forgetting,” was first established by Ebbinghaus in 1885 and has been extremely well documented in the ensuing years (Anderson & Schooler, 1991; Ebbinghaus, 1885, 1913; Wixted & Ebbesen, 1991; Wixted, 2022).

If more experts believe that eyewitness confidence does not predict accuracy than believe in a well-established principle of forgetting, then it seems fair to say that the issue was considered to be settled science. This, in turn, would explain why this idea made its way into multiple amicus briefs filed by the American Psychological Association. It would also explain why it made its way into textbooks, which is presumably why, as recently as of 2019, a remarkable 82% of surveyed undergraduates agreed with the following statement: “An eyewitness’s confidence is never a good predictor of his or her identification accuracy” (Brewin et al., 2019, emphasis added).

Part 2: The Evolution of Eyewitness Memory Research

Things have changed, and to appreciate why, it is important to first understand the details of a lineup procedure. These days, law enforcement agencies predominantly use photo lineups (not live lineups) to test eyewitness memory. Photo lineups consist of one suspect—whose guilt or innocence is to be determined—alongside a set number of known-to-be-innocent fillers, typically five in the U.S. The eyewitness can either identify an individual from the lineup (the suspect or one of the fillers) or they can reject the lineup altogether if they believe the perpetrator is not present. Longstanding recommendations for administering a fair lineup procedure include using fillers who match the witness's description of the perpetrator, implementing a double-blind procedure (i.e., neither the witness nor the lineup administrator is told who the suspect is), warning the witness that the perpetrator may or may not be in the lineup, and collecting an immediate confidence statement (Wells et al., 1998; 2020).

In actual criminal investigations, it is impossible to know whether a suspect is guilty prior to a lineup. However, this can be easily manipulated in laboratory settings by simply changing the suspect presented to the witness. When the suspect is the actual perpetrator, the lineup is called a "target-present" (TP) lineup. When the suspect is an innocent person who bears a passing resemblance to the actual perpetrator—this is considered a "target-absent" (TA) lineup. In a typical lab study, each participant observes a mock crime and then is tested only once using a TP or TA lineup. Thus, the data consist of suspect IDs, filler IDs, and lineup rejections from TP lineups and suspect IDs, filler IDs, and lineup rejections from TA lineups. Each lineup decision is either correct or incorrect (0 or 1) and is accompanied by a confidence rating. The correlation between confidence and accuracy is computed from these data.

As noted earlier, even using a proper lineup procedure and testing uncontaminated memory, research showed that the correlation between confidence and accuracy was weak. However, Juslin et al. (1996) was the first to explain why using point-biserial correlation as an indicator of confidence-accuracy relationship is problematic. They argued that this measure is not only less than optimal, but it is also potentially misleading (possibly extremely so). Instead, Juslin et al. recommended an alternative method known as calibration analysis, which quantifies the correspondence between confidence measured on a hundred-point scale and percent-correct accuracy, which is also measured on a hundred-point scale. Perfect calibration exists when eyewitnesses who express 100% confidence are 100% accurate, those who express 75% confidence are 75% accurate, and so forth. Critically, the point-biserial correlation coefficient can be misleadingly close to 0 even when confidence is *perfectly* calibrated to accuracy (Juslin et al., 1996). Subsequent studies corroborated the value of calibration analysis. For example, using that approach, Brewer & Wells (2006) reported a high confidence-accuracy relationship, particularly among "choosers" (i.e., those who made a positive ID of the suspect of a filler).

However, even this undeniably improved method for assessing the confidence-accuracy relationship ultimately turned out to be less than optimal as well. The calibration approach is perfectly fine for an identification procedure like a showup, which involves a single suspect (innocent or guilty). In a showup, to compute accuracy for a given level of confidence (e.g., high-confidence decisions made with 90-100% confidence), one simply uses the formula 100% × $\frac{GS_{high}}{GS_{high}+IS_{high}}$, where GS_{high} represents the number of correct identifications of the guilty suspect made with high confidence (i.e., 90-100%), and IS_{high} represents the number of incorrect identifications of the innocent suspect made with high confidence. Similar values can be computed for the remaining levels of confidence, and then the computed accuracy scores can then be plotted as a function of confidence to create a calibration curve.

Unlike a showup, a lineup has fillers, and intuition suggests that filler IDs should somehow be considered in the calibration score as well. Whether a filler ID is made from a TP lineup or a TA lineup, it is incorrect, which raises a question: How should these additional incorrect decisions be factored into the calibration equation?

The calibration curve of most interest is the one that represents the performance of "choosers" (i.e., those who positively identify someone). For a positive ID, the only correct response is an identification of the guilty suspect. Thus, the numerator of the calibration equation consists only of such IDs made with a certain level of confidence, whereas the denominator consists of all identifications (correct or incorrect) made with that same level of confidence. Thus, it might make sense to add incorrect filler identifications (i.e., the number of filler IDs from TP lineups and the number of filler IDs from TA lineups) to the denominator of the equation presented above for the showup. In practice, it is sometimes done that way (e.g., Brewer et al., 2002), but more often errors from TP lineups are excluded on the grounds that they are known to be innocent (e.g., Brewer & Wells, 2006). However, the fillers in TA lineups are also known to be innocent and could be similarly excluded. But they never are.

The fact that TA filler IDs are never excluded from calibration analysis reflects a methodological feature commonly used in lineup studies. Specifically, unlike in real police lineups, in lab studies, TA lineups often have no designated innocent suspect (i.e., the lineup consists only of filler photos) and thus have no predefined set of innocent fillers distinct from the suspect to exclude. It makes sense to create TA lineups in this way because when a real lineup contains an innocent suspect, from the witness's perspective, the lineup consists of nothing but fillers (i.e., nothing but people who match the witness's description of the perpetrator but who did not commit the crime). An innocent suspect is special only from the perspective of the police because they mistakenly believe he might be guilty.

Because of this design feature, should filler IDs from TA lineups be included in the calibration score even while excluding filler IDs from TP lineups? This is not really a sound approach. If one is going to compute calibration, and if TA filler IDs are going to be included in the calculation as errors, then TP filler IDs should also be included as errors. As recently as 2014, an APA amicus brief used this more reasonable approach and cited calibration research as demonstrating the following:

> "Importantly, error rates can be high even among the most confident witnesses. Researchers have performed studies that track, in addition to identification accuracy, the subjects' estimates of their confidence in their identifications. 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 confident in the accuracy of their identifications…This confirms that many witnesses are over-confident in their identification decisions" (American Psychological Association, 2014b, pp. 17–18)."

Thus, even though calibration research made it clear that eyewitness confidence was more informative than previously assumed, the longstanding message from psychological science to the courts remained essentially the same: eyewitness identifications are unreliable even when made from a pristine lineup, even before memory was contaminated, and even when confidence was high. But that message was about to change.

Post-2010 marked a significant shift in the field of eyewitness memory research because it is when basic memory research in the field of cognitive psychology began to cross-pollinate with its applied counterpart in the field of eyewitness identification. Cognitive psychologists took a different approach to studying eyewitness identification by framing as recognition tests as they are often framed in basic research, namely, as a signal detection task (Mickes et al. 2012, Wixted & Mickes, 2014; Wixted, Vul, et al., 2018). Signal-detection theory is a well-established framework to conceptualize how individuals discriminate between meaningful signals against random noise (Green & Swets, 19666; Tanner, 1961). As applied to eyewitness identification, and in its simplest form, signal detection theory holds that the memory signals generated by guilty suspects across lineups are normally distributed. On average, guilty suspects generate a relatively strong memory signal because the guilty face matches the face of the perpetrator stored in the witness's memory. The strength of the memory signals generated by innocent suspects (and innocent fillers) are also normally distributed across lineups, but because these faces do not match memory of the perpetrator, the memory-match signal is weaker, on average. For a given lineup, one face will generate the strongest memory signal, and only that face is a candidate for being identified. If the memory strength of that face exceeds a decision criterion, then it is identified (with confidence increasing the more the signal exceeds the criterion); if not, the lineup is rejected.

It is well known that the basic signal detection model of recognition memory inherently predicts a strong confidence–accuracy relationship, so it initially seemed like a poor fit to eyewitness identification research (Wixted, 2020). But the problem was not with signal detection theory. Instead, the problem was with how the confidence-accuracy relationship was being assessed in terms of correlation or calibration.

Mickes (2015), drawing on Bayesian principles in conjunction with signal detection theory, introduced the Confidence-Accuracy Characteristic (CAC) as an enhancement over calibration analysis methods when a lineup is used. In a legal context, the question of interest is usually how reliable a *suspect* ID is, given that the ID was already made. By contrast, witnesses who initially make filler identifications and lineup rejections should not make it to the witness stand to positively identify the suspect (although, as discussed later, they sometimes do). Those who initially identify the suspect are usually the ones who end up testifying about their positive identification. The CAC approach therefore calculates the accuracy of suspect identifications per se at each level of confidence, providing a more relevant metric. In other words, CAC analysis answers the following question: given that the eyewitness identified a suspect from the lineup with a certain level of confidence, how likely is that identification to be correct? Signal detection theory predicts a strong confidence-accuracy relationship for *this* measure (i.e., the one judges and juries care about).

In CAC analysis, the formula for computing accuracy is the same one presented earlier for computing calibration when a showup procedure is used. In that formula, only correct and incorrect suspect IDs are included (because there are no filler IDs when a showup is used). The same formula should be used even when fillers are present (i.e., even when a lineup is used). In other words, the information value of filler IDs from TA and TP lineups should be assessed by analyzing them separately from suspect IDs. CAC analysis keeps the focus squarely on correct and incorrect suspect IDs made with different levels of confidence. But how does one compute innocent suspect IDs from TA lineups in the common lab design in which these lineups consist of nothing but fillers (i.e., no designated innocent suspect)? In that case, incorrect innocent suspect IDs can be estimated by simply dividing the number of TA filler IDs by lineup size. In essence, one way or another, CAC analysis excludes filler IDs from both TP and TA lineups.

Wixted and Wells (2017) took CAC analysis and retroactively applied it to data from 15 simulated crime studies that previously employed point-biserial correlation and calibration analyses. Their reanalysis showed a new pattern: suspect IDs made with high confidence were highly accurate (well above 90% correct), whereas the ones associated with low confidence were not much better than chance (Wixted et al., 2016). In other words, far from being uninformative, on an initial, properly conducted test of uncontaminated memory, an eyewitness's expressed level of confidence could scarcely be more informative than it is. Critically, this is the very definition of reliability. Thus, the longstanding and almost universal verdict concerned with the unreliability of eyewitness memory was turned on its head.

Remarkably, this strong confidence-accuracy relationship was also present in real-world data involving actual eyewitnesses in investigations conducted by the Houston Police Department (Wixted et al., 2016). This study adhered to recommended pristine lineup procedures, so the findings were unlikely to be artifacts of unfair lineups or police steering witnesses to pick suspects. The key finding was that when witnesses picked a filler, they often did so with low confidence, but when they picked a suspect, they typically did so with high confidence. Given that these photo lineups were conducted early in a police investigation (thereby minimizing the chances that memory had already been contaminated) and in such a way that a witness could only land on a suspect at above-chance levels using their memory of the perpetrator, the results were entirely consistent with the new understanding of eyewitness reliability based on lab studies.

Subsequent research also using real-world lineups confirmed this robust confidence-accuracy relationship (Quigley-McBride & Wells, 2023) and additionally underscored another variable that was found to be associated with accuracy: response time. Basically, as lab-based research had long shown, IDs made quickly tend to also be made with high confidence, and the same is true of real-world identifications from police lineups (Seale-Carlisle et al., 2019). Together, confidence and decision time were dubbed “reflector variables” (i.e., variables that reflect identification accuracy). Decision time and confidence tend to be inversely correlated (the higher the confidence, the faster the decision time), so they are somewhat redundant measures. However, it is important to also consider decision time because on the rare occasions in which a high-confidence identification is made slowly (e.g., after 60 seconds), accuracy tends to be noticeably lower compared to the usual case in which a high-confidence identification is made quickly (e.g., less than 10 seconds).

With compelling empirical evidence now spanning both simulated and real-world settings, the 2010s emerged as a period that saw cognitive psychologists build a cogent, theoretically grounded case for the predictive value of confidence and the reliability of eyewitness memory. Research in the field is now using detailed signal-detection modeling to dive deeper into the mechanisms that underlie the process of eyewitness identification (Wixted, Vul, et al., 2018; Shen et al., 2023; Fitousi, 2023).

It is worth noting that the confidence-accuracy relationship that we have discussed thus far does not conflict with earlier research attesting to the frailties of human memory. For instance, the wrongful convictions overturned by the Innocence Project make it clear that an eyewitness's testimony at trial is often tainted by post-identification misinformation. Precisely because memory is malleable, the strong confidence-accuracy relationship can only be observed in *initial* IDs (prior to memory contamination) obtained under pristine testing conditions. Therefore, rather than being at odds, these two perspectives can be seen as complementary factors of a complex issue, each contributing insight into the broader understanding of eyewitness memory and its role in the legal system.

Part 3: The Mechanisms of Eyewitness Memory

So far, we have briefly considered the role of signal detection theory in helping to understand the confidence-accuracy relationship in eyewitness identification. However, there is much additional work from the basic science-of-memory literature that needs to be considered in order to fully appreciate why everything we just explained with regard to the impressive confidence-accuracy relationship applies to the *first* test of an eyewitness's memory for a suspect. As a general rule, no later test provides more reliable information than the first because (1) the first test minimizes the chances of testing contaminated memory and (2) the test itself contaminates the witness's memory for the suspect. Thus, from a scientific perspective, the first test is of special importance. By contrast, from a legal perspective, the last test (at trial) is of special importance. The last test is of special importance from a legal perspective because the witness (1) is under oath and (2) can be cross examined by opposing counsel. However, the test that occurs at the criminal trial provides the least reliable information because it maximizes the chances of testing contaminated memory. Placing the witness under oath and allowing the witness to be cross examined cannot change that unfortunate fact. Basically, the choice is between reliable information obtained when the witness is not under oath and cannot be cross examined vs. unreliable information obtained when the witness is under oath and can be cross examined.

In the personal experience of the senior author of this chapter, prosecutors sometimes worry that putting an emphasis on the initial identification test might simply create a loophole for the benefit of the perpetrator. For example, imagine that on the initial test, the witness rejects a lineup containing a recognizable photo of the suspect. However, the police nonetheless remain convinced that the suspect is the perpetrator and proceed to test the witness again, perhaps using a live lineup this time on the theory that it provides a better test of memory. Now, the witness might identify the suspect for the first time, and the police will likely conclude that, on the second test, the witness "found" the memory of the perpetrator that they failed to find on the first test. After all, everyone has had the experience of failing to remember a detail on the first try (e.g., the name of a recent acquaintance) and then successfully remembering it days, weeks, or even months later. Perhaps something similar is happening when the witness finally recognizes the suspect that the police already believe may have committed the crime.

However, that is not what is happening, and believing otherwise reflects an understandable yet catastrophic lack of expertise on the part of judges and jurors. It is certainly true that certain kinds of memory, specifically recollection memory (e.g., recalling someone's name), can initially fail but succeed later—a phenomenon technically known as "reminiscence" (Gilbert & Fisher, 2006). Reminiscence can lead to overall memory performance improving as a function of repeated testing for a single encoding event (e.g., repeatedly testing memory for the once-presented list of words). This overall improvement in performance is known as "hypermnesia." However, it is crucial to understand that this phenomenon applies to the recollection of details but does not ordinarily apply to a recognition test (Payne & Roediger, 1987), and a lineup test is a recognition test. Why not?

Recollection tests ask for information that is not physically present during the test—e.g., "What was the perpetrator wearing?" This retrieval cue prompts the eyewitness to conduct a probabilistic search of memory to retrieve the relevant detail. Because of its probabilistic nature, a failure in the initial test does not distort the stored memory, so if the attempt to find that information on a later try, it is not inherently problematic. Analogously, you can fail to find your keys when first searching for them, but if you find them when you search again later, the keys are still exactly as they were before (having been unaffected by the initial failed search).

By contrast, recognition memory tests ask about a stimulus that is physically present. In a lineup, for example, the question is: "Do you see the face of the person who committed the crime?" The nature of this test involves picking out the option that elicits the strongest familiarity signal (or rejecting the lineup if even that face is not sufficiently familiar). The face familiarity signal is immediately available because it is not the result of a search process. Instead, it is the result of a *matching* process.

To appreciate this key issue, consider Mandler's classic "Butcher on the Bus" anecdote (Mandler, 1980). This anecdote describes an experience of encountering a man on a bus whose face is so familiar that it calls for a search of memory to determine the source of familiarity. The source of the familiarity signal is not physically present and therefore requires a search of memory to find the original encoding context ("Where do I know him from?"). However, the familiarity signal itself—that is, the familiarity signal associated with the physically-present face—is not like that. The familiarity signal is experienced promptly upon viewing the face because no search of memory is required.

In a police lineup, the face is physically present (no search required) and so is the relevant context in which the face might have been previously encountered because the police ask the witness who committed the crime (again, no search required to find the relevant context where the face might have been originally encoded). Because everything that one might have to search for in memory is already available, if a face in the lineup is not immediately familiar, no amount of searching memory or analyzing facial features or comparing and contrasting the faces ("it is between these two") will change that fact. Instead, the lack of a strong and promptly experienced familiarity signal upon viewing the faces in the lineup means that none were encoded into the witness's memory. None of the faces are familiar either because they are all innocent (including the suspect) or because the suspect is guilty, but the witness did not get a good look at the perpetrator at the time of the crime.

Unfortunately, witnesses do not understand this principle, and neither do the police. Thus, it is not uncommon for a witness to spend *minutes* searching memory in a hopeless effort to find the face familiarity signal somewhere in their brain. Jennifer Thompson's initial identification of Ronald Cotton, for example, came after almost 5 minutes of analyzing the faces in the lineup.

But 10 or 20 seconds into the lineup, it was already clear that no one matched the memory of the perpetrator encoded in her brain (including Rondald Cotton). Had everyone understood what we know now—namely, that on this crucial initial test of uncontaminated memory, Jennifer Thompson made an ID with very low confidence—Cotton may never have been wrongfully convicted.

Critically, there are no second chances to test a witness's uncontaminated memory for a suspect, and this is true no matter how perfect or how suggestive the initial lineup procedure was and no matter what the outcome of the initial test was (a suspect ID, a filler ID, or no ID). One way to appreciate why this is true is to consider the "memory for foils" paradigm (Jacoby et al., 2005) in which participants first study of list of words followed a recognition test in which they are asked to distinguish between words that appears on the list (targets) and words that did not (foils). Later, the participants are given an unexpected recognition test for the foils (now targets), which are intermixed with another set of novel foils. Note that when taking the first recognition test, the participants were only making recognition decisions, not attempting to memorize anything. Moreover, they correctly rejected most of the foils that would appear on the subsequent surprise recognition test. After all, there is nothing suggestive about a memory test like this. Yet on the surprise test that followed (testing memory for the recently seen foils), their memory performance was as good as it was for the targets on the first test. The clear implication is that testing memory by means of recognition creates a memory of the tested items, including the foils.

Somehow, it has not been sufficiently appreciated that the same memory principle applies when memory is tested by recognition using a lineup. In a lineup test, an innocent suspect is a foil. Thus, there are no second chances to test uncontaminated memory for a suspect because the act of taking the test and evaluating the options creates a representation of the faces in the lineup (including the suspect's face). Even worse, these memory records are being created at a time when the witness is thinking about the crime, thereby associating the face with the crime. Thus, if any one of these faces appears in a subsequent test (as only the innocent suspect's face will), it will seem more familiar than would otherwise be the case and the source of that familiarity signal may be mistakenly attributed to the crime. Pretending that such contamination has not happened and that the witness at trial can instead find an untainted, independent source of their memory for what happened on the day of the crime is a game that results in tragic consequences.

If the witness recognizes a face on a second test after failing to recognize it with high confidence on the first test, it is not an instance of reminiscence but is instead a reflection of memory contamination. *This* is how witnesses come to confidently misidentify an innocent defendant in front of a jury, leading to a wrongful conviction, not because eyewitness memory is unreliable. Witnesses are reliable on the first test, and it is not their fault that no one listens to what they have to say at that time. Instead, the rules of evidence at a criminal trial are such that judges and juries are almost certain to consider what the witness has to say at trial, long after the witness's memory has been contaminated. The contamination begins with the first lineup test and is later exacerbated by many subsequent events, such as confirming feedback from police or prosecutors (Wells & Bradfield, 1998). And all the while, as the police keep the focus on the innocent suspect (right up until the criminal trial a year or two later), the memory of the actual once-seen perpetrator in the brain of the eyewitness is gradually fading away.

Prisoners who are now serving long sentences and whose cases illustrate this problem are not hard to find. In fact, this may be the most important implication of the new understanding according to which there is only one chance to a witness's uncontaminated memory for a suspect and on that initial test, eyewitnesses provide reliable information. For a field that has, for decades, emphasized the inherent unreliability of eyewitness memory, the idea that memory actually tends to be reliable on the first test is disconcerting. The fear is that witnesses will do on the first test what they often do on the last test at trial, namely, misidentify an innocent suspect with high confidence. Contrary to that understandable concern, the opposite seems to be true.

Garrett (2011) analyzed trial testimony from 161 DNA exoneration cases and found that in 57% of those cases, there was testimony about what the eyewitness did on the first test (92 cases in all). These 92 witnesses all testified with high confidence at trial, of course, but how many did so on the first test involving the suspect who would later be wrongfully convicted? Not a single one did so. And why would they? That is, why would a witness immediately and confidently misidentify the face of an innocent suspect that does not match their fresh memory of the perpetrator? Instead, and understandably, all of these witnesses identified the suspect with low confidence or picked a filler or rejected the lineup. Some said they did not even see the perpetrator in the first place. Garrett's (2011) findings, perhaps more than any other, underscore the importance of keeping the focus on the first test (i.e., the test that provides reliable information). They also underscore the point from lab studies that on the first test, eyewitnesses are not unreliable. Instead, it is the criminal justice system that is unreliable because it does not listen to what the witness says on the first test and then repeatedly tests the witness's' contaminated memory to win a conviction.

Once the focus is placed on the first test, it becomes clear that there are many cases just like the DNA exoneration cases but for which no DNA evidence is available. Thus, for these unfortunate prisoners, no DNA exoneration will ever happen. In these cases, the apparent evidence of guilt that led to the conviction (namely, the high-confidence testimony of eyewitnesses at the criminal trial), properly understood, actually points in the opposite direction (i.e., on the initial test, the exact same eyewitnesses provided clear evidence of innocence). We end this chapter by considering one case that illustrates the importance of the story we have told here. However, this is by no means the only case.

Part 4: Case Illustration

In 2009, James Joseph Garner was at a bar in Denver Colorado when a gunman opened fire, injuring brothers Arturo, Christian, and Roberto Adam-Diaz (People of the State of Colorado v. James Joseph Garner, 2015). Immediately after the shooting, witness #1 (Roberto) told police that he had not clearly seen the shooter. Witness #2 (Arturo) did see the shooter described him as a man who was 27 years old and 5 feet, 2 inches tall, but Garner was 9 years older and 6 inches taller than that. Witness #3 (Christian) also saw the gunman and provided the most detailed description, telling the police that he was bald, wearing no glasses, with no facial hair, and with a tattoo on his head. Unlike the gunman Christian described, Garner had short black hair, glasses, facial hair, and no such tattoos. Thus, it seems clear that Christian described a completely different person. On an initial test of uncontaminated memory (but not on later tests of contaminated memory), witness descriptions tend to be highly reliable (e.g., Wixted et al., 2018), so, already at this early stage of the investigation, the witness descriptions would seem to rule out Garner as being the shooter.

The police nevertheless continued to suspect Garner and placed his photo in an array along with five similar fillers. This was the key test, the one that provides the most reliable information because it minimizes the chances of testing contaminated memory. On this key test, *none of the three witnesses selected Garner*. The fact that all three witnesses looked at Garner's face and did not identify him provides compelling evidence that Garner is innocent (on top of the evidence of innocence provided by the witness descriptions of the perpetrator). But this first test also contaminated the memories of the three witnesses by creating a memory of Garner's face at the time they were thinking about the crime. Many more opportunities for memory contamination would occur in the subsequent *three years* leading up to Garner's criminal trial.

Of course, at the trial, all three witnesses confidently identified Garner. You might think that, upon hearing that the witnesses were unable to identify Garner on the initial test, the jury would scoff at the high-confidence IDs that occurred at trial and return a verdict of "not guilty." That is, you might think that their prior inconsistent statements would impeach the credibility of the witnesses. However, it would be a serious mistake to think along these lines. Recall, for example, the 92 cases analyzed by Garrett (2011) in which the inconclusive ID came up at trial and the jury still found the defendant guilty every time. A fundamental problem is that when a witness's memory changes from inconclusive to conclusive, a non-science-based explanation for the change will almost always be offered (either by the witness or by the prosecutor) and juries will almost always uncritically accept that explanation (Yilmaz et al., submitted).

In the Garner trial, the prosecutor argued that the witnesses failed to identify Garner from mere photos but recognized him 3 years later at trial because now (at trial) the identification procedure was an in-person test. One almost has to admire this brazen move given that police departments use photo lineups (not live lineups) almost exclusively these days. According to the prosecutor's logic in the Garner case, in-court IDs following an initial failure to identify the suspect from a photo lineup is not a recipe for a wrongful conviction but instead serves the cause of justice. However, the relevant science says it is the other way around.

Contrary to the prosecutor's assertion, multiple studies have found that live lineups yield no better performance than photo lineups. Fitzgerald et al. (2018) reviewed the relevant literature on the live lineup superiority hypothesis and succinctly summarized their findings as follows: "In spite of our efforts to be as charitable as possible to the live superiority hypothesis, we found little reason to support it" (p. 321). In a subsequent study, Rubínová et al. (2021) conducted a large-scale investigation of this issue, and the title of their paper makes their findings clear: "Live presentation for eyewitness identification is not superior to photo or video presentation." Thus, the prosecutor was just making up a believable story for the jury to embrace, which, perhaps surprisingly, prosecutors are entitled to do.

What other explanation might explain why all three witnesses failed to identify Garner on the all-important first test but did so with ultimate confidence in front of the jury at the criminal trial? The alternative explanation considered at Garner's trial is that the witnesses might be lying, but the prosecutor quite reasonably rejected that theory:

> "Is it possible that all three of these guys come in here and lie to you and tell you that they're 100 percent sure he's the shooter and that they're all three willing to send an innocent person to get convicted of this? I would submit to you that its not even possible, but it's certainly not reasonable" (Colorado Court of Appeals, People of the State of Colorado vs. James Joseph Garner, 2015, p. 27).

Faced with these two possible explanations for why the witnesses were identifying Garner now despite not having done so previously—namely, they had better memory for live faces seen now than for photos seen 3 years ago vs. they are all simply lying—the jury apparently accepted the former one. Garner was found guilty and sentenced to 35 years in prison.

But there is a third explanation that fits the facts much better than the other two theories, one that was never considered at the criminal trial or during his subsequent appeals. It is a theory that is based on a new scientific consensus that emerged in 2020: testing a witness's memory for a suspect unintentionally contaminates the witness's memory for that suspect (Wells et al., 2020; Wixted et al., 2021). A few exceptions notwithstanding (e.g., when the witnesses were lying not at trial but on the first test), the contaminating effect of the first test renders all later tests less reliable than the first. On the first and only test of uncontaminated memory, the three witnesses clearly exonerated James Garner.

In this case, and in many others just like it, it is not the witnesses who were unreliable. Misdiagnosing the problem of wrongful convictions as being attributable to the unreliability of eyewitness memory distracts from solving the problem. The problem is that the legal system is designed to almost ensure a focus on the least reliable test of a witness's memory while simultaneously disregarding the most reliable one. To solve that problem, the legal system must find a way to keep the focus on the reliable first test and take it off of the unreliable last test at trial.

Although the field has established the reliability of eyewitness memory on an initial, properly conducted identification test, almost everything about the judicial system in the U.S. prioritizes what the eyewitness has to say on the last test, where the witness is under oath and available for cross-examination in the presence of a judge. If a witness rejected an initial lineup containing a good photo of the suspect (now defendant), it can be difficult to get that lineup rejection admitted as substantive evidence of innocence. This is especially true if, at trial, the witness mistakenly remembers having always been confident that the defendant is the perpetrator, in which case the police report indicating a lineup rejection is hearsay (a matter asserted out of court while not under oath). Going forward, it seems essential to educate various actors in the legal system (judges, prosecutors, defense attorneys) about the new science-based consensus on the importance of focusing on the first test of a witness's memory.

References

American Psychological Association. (2014a). *Brief for Amicus Curiae, Commonwealth v. Walker*. Retrieved from: http://www.apa.org/about/offices/ogc/amicus/walker.aspx

American Psychological Association. (2014b). *Brief for Amicus Curiae, Commonwealth of Massachusetts v. Gomes*. Retrieved from: https://www.apa.org/about/offices/ogc/amicus/gomes-johnson

American Psychological Association. (2016). *Brief for Amicus Curiae, Payne v. Commonwealth of Virginia*. Retrieved from http://www.apa.org/about/offices/ogc/amicus/payne.pdf

American Psychological Association. (2019). *Brief for Amicus Curiae, Watson v. Commonwealth of Virginia*. Retrieved from https://www.apa.org/about/offices/ogc/amicus/watson-virginia.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2(6)*, 396–408.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2020). *Memory*. Psychology Press.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology*, *72*(4), 691.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8*(1), 44–56. https://doi.org/10.1037/1076-898X.8.1.44

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

Brewin, C. R., Li, H., Ntarantana, V., Unsworth, C., & McNeilis, J. (2019). Is the public understanding of memory prone to widespread “myths”? *Journal of Experimental Psychology: General, 148*, 2245–2257.

Deffenbacher, K. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*, 243-260.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). Teachers College Press. (Original work published 1885).

Ebbinghaus, H. (1885). Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Leipzig: Duncker & Humblot.

Ebbinghaus, H. (1913). Memory: A Contribution to Experimental Psychology. New York: Teachers College, Columbia University.

Eyewitness misidentification. Innocence Project. (2023, May 1). https://innocenceproject.org/eyewitness-misidentification/

Fitousi, D. (2023). A signal detection–based confidence–similarity model of face matching. *Psychological Review*.

Fitzgerald, R. J., Price, H. L., & Valentine, T. (2018). Eyewitness identification: Live, photo, and video lineups.*Psychology, Public Policy, and Law, 24*(3), 307–325. https://doi.org/10.1037/law0000164

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

Garrett, B. L. (2012). Eyewitnesses and exclusion. *Vanderbilt Law Review*, *65*, 451-506.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969-2012). New York: Wiley.

Gilbert, J. A., & Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology*, *20*(6), 723-739.

Innocence Project (2023). Retrieved from https://www.innocenceproject.org/exonerate/

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: depth of retrieval. *Psychonomic Bulletin & Review, 12(5)*, 852–857. https://doi.org/10.3758/bf03196776

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1304-1316.

Kassin, S. M., Ellsworth, P., & Smith, V. L. (1989). The "general acceptance" of psychological research on eyewitness testimony. American Psychologist, 44, 1089-1098.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56*(5), 405–416. https://doi.org/10.1037/0003-066X.56.5.405

Leippe, M., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. *In The Handbook of Eyewitness Psychology: Volume II,* pp. 391-440. Psychology Press.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of experimental psychology: Human learning and memory*, *4*(1), 19.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*(5), 585-589.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*(12), 720-725.

Luus, C. A., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology*, *79*(5), 714.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252– 271.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*(1), 1.

Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *4*(2), 93.

Mickes, L., Flowe, H. D. & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous and sequential lineups. *Journal of Experimental Psychology: Applied, 18*, 361-376.

Payne, D. G., & Roediger, H. L. (1987). Hypermnesia occurs in recall but not in recognition. *American Journal of Psychology, 100*, 145–165.

People of the State of Colorado v. James Joseph Garner (2015). Colo. App. 12CA2540. https://cases.justia.com/colorado/court-of-appeals/2015-12ca2540.pdf?ts=1450368352

Quigley-McBride, A., & Wells, G. L. (2023). Eyewitness confidence and decision time reflect identification accuracy in actual police lineups. *Law and Human Behavior*.

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.*Journal of Applied Research in Memory and Cognition, 10*(1), 167–176. https://doi.org/10.1016/j.jarmac.2020.08.009

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

Schacter, D. L. (2022). Memory sins in applied settings: What kind of progress? *Journal of Applied Research in Memory and Cognition*, *11*(4), 445.

Seale-Carlisle, T. M., Colloff, M. F., Flowe, H. D., Wells, W., Wixted, J. T., & Mickes, L. (2019). Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world. *Journal of Applied Research in Memory and Cognition*, *8*(4), 420-428.

Shaw, J., & Porter, S. (2015). Constructing rich false memories of committing crime. *Psychological science*, *26*(3), 291-301.

Shen, K. J., Colloff, M. F., Vul, E., Wilson, B. M., & Wixted, J. T. (2023). Modeling face similarity in police lineups. *Psychological Review, 130(2)*, 432–461. https://doi.org/10.1037/rev0000408.

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. *Psychological Bulletin*, *118*(3), 315.

Tanner, W. P. (1961). Physiological implications of psychophysical data. *Annuals of the New York Academy of Science, 89*, 752-765.

Thompson-Cannino, J., Cotton, R., & Torneo, E. (2009). *Picking cotton: Our memoir of injustice and redemption*. Macmillan.

Wade, K. A., Garry, M., Don Read, J., & Lindsay, D. S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic bulletin & review*, *9*(3), 597-603.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology*, *83*, 360-376.

Wells, G. L., 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 human behavior*, *44*(1), 3.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). New York, NY: Cambridge University Press.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human behavior*, *22*(6), 603.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of experimental psychology: learning, memory, and cognition*, *46*(2), 201.

Wixted, J. T. (2022). Absolute vs. relative forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*, 1775-1786.

Wixted, J. T., & Ebbesen, E. (1991). On the form of forgetting. *Psychological Science, 2*, 409-415.

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

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences, 113(2)*, 304-309.

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the Reliability of Eyewitness Memory. *Perspectives on Psychological Science, 13*, 324-335.

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive psychology*, *105*, 81-114.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*(1), 10-65.

Wixted, J. T., Wells, G. L., Loftus, E. F., & Garrett, B. L. (2021). Test a witness's memory for a suspect only once. *Psychological Science in the Public Interest, 22*(1_suppl), 1S-18S. https://doi.org/10.1177/15291006211026259

Zaragoza, M. S., McCloskey, M., & Jamis, M. (1987). Misleading postevent information and recall of the original event: Further evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 36.