Eyewitness Memory is Reliable, but the Criminal Justice System is not

John T. Wixted

University of California, San Diego

Laura Mickes

University of Bristol

Author Note

John T. Wixted, Department of Psychology, University of California, San Diego. Laura Mickes, School of Psychological Science, University of Bristol

Correspondence concerning this article should be addressed to John Wixted (jwixted@ucsd.edu).

Abstract

The reliability of any type of forensic evidence (e.g., forensic DNA) is assessed by testing its information value when it is not contaminated and is properly tested. Assessing the reliability of forensic *memory* evidence should be no exception to that rule. Unfortunately, testing a witness's memory irretrievably contaminates it. Thus, only the first (properly conducted) test is relevant to the question of whether eyewitness memory is reliable. With few exceptions, the results of studies conducted in the lab and in the real world show that confidence is highly predictive of accuracy on the *first* test, and high-confidence often implies high accuracy. The fact that many eyewitnesses are known to have made high-confidence misidentifications in the courtroom has cemented the almost universal impression that eyewitness memory is unreliable. However, it is the criminal justice system that is guilty of using contaminated memory evidence (relying on the *last* memory test, in court) in conjunction with an improper testing procedure (namely, a courtroom showup) to unwittingly win convictions of the innocent. That mistake should no longer be blamed on the unreliability of eyewitness memory.

Key words: Eyewitness Identification; Wrongful Convictions; Malleability of Memory

Eyewitness Memory is Reliable, but the Criminal Justice System is not

"There is almost nothing more convincing than a live human being who takes the stand, points a finger at the defendant, and says 'That's the one!'" -- William J. Brennan

Jurors presumably find confident witnesses convincing because, in everyday life, confidence is predictive of accuracy on a wide range of tasks involving memory and perception. Indeed, for such tasks, computational models designed to account for the strong relationship between (1) confidence, (2) reaction time, and (3) accuracy have proliferated in recent years, both in psychology (e.g., Moran et al., 2015; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2013) and in neuroscience (e.g., Kiani & Shadlen, 2009; Kiani et al., 2014). Even rats appear to show a strong confidence-accuracy relationship, with high confidence associated with high accuracy (Kepecs et al., 2008). Yet ~70% of the wrongful convictions overturned by DNA evidence were caused, at least in part, by confident eyewitnesses misidentifying innocent defendants in a court of law (Innocence Project 2021). Is eyewitness memory a special case, where the usual relationship between confidence and accuracy does not apply for some reason?

We apparently teach our students to think so. In a recent survey, Brewin et al. (2019) found that 82% of undergraduates agreed with the following statement: "An eyewitness's confidence is *never* a good predictor of his or her identification accuracy" (emphasis added). Contrary to that belief, confidence is strongly related to accuracy on a properly conducted test of uncontaminated eyewitness memory, whether memory is tested by recall (e.g., an interview involving open-ended questions) or recognition (e.g., a fair photo lineup). This is true of data both from the lab and from the real world. The key to understanding the disconnect between what students believe vs. objective reality is to recognize that the iconic cases of misidentifications leading to wrongful convictions come from the *last* test of eyewitness memory, conducted in court, long after memory was contaminated. The focus should instead be

on the *first* test, which unarguably minimizes the chances of testing contaminated memory (Wixted et al., 2015).

Consider, as an analogy, DNA evidence. Imagine that whenever a forensic DNA test failed to implicate a suspect in a crime (i.e., the test result came back either "inconclusive" or "no match"), the police routinely asked the suspect to touch the evidence and then had the forensic laboratory run the test a second time (now obtaining a perfect match). Further imagine that a prosecutor used the results of that second test to convince a jury that the suspect is guilty. Contrary to Supreme Court Justice William Brennan's famous assertion, this test result might be even more convincing to a jury than a confident eyewitness.

In this hypothetical scenario, would the correct diagnosis of the problem be that forensic DNA evidence is unreliable, or would a better diagnosis be that the criminal justice system is guilty of contaminating the forensic DNA evidence and then using it to convict defendants, innocent or guilty? Most would presumably agree with the latter diagnosis. This analogy is relevant to eyewitness memory because the very act of testing memory contaminates memory (Wixted et al., in press). Therefore, any later test of the same content—that is, the same already-recollected details (e.g., Odinot, Wolters, & Lavender, 2009) or the same suspect who was already viewed in an earlier lineup (e.g., Steblay et al., 2013)—will result in confidence inflation and therefore constitutes a test of contaminated memory evidence.[1]

If testing memory contaminates memory, then the question of how reliable eyewitness memory is pertains *only* to an initial test. As we consider next, the evidence overwhelmingly

[1] A second interview of a witness can usefully elicit recollection of new details (e.g., Odinot, Memon, La Rooy, & Millen, 2013), but it is the first interview that provides the relevant information about already-recollected details.

shows that, when properly tested, eyewitness memory is reliable in that confidence is highly predictive of accuracy, with high confidence memories often being impressively accurate.

Memory tested using an interview (recall of details)

Consider, first, studies in which information was elicited using an initial interview involving open-ended, non-leading questions. Roberts and Higham (2002) conducted the first study that we are aware of examining whether confidence predicts eyewitness recall accuracy. Participants watched a videotape of a simulated robbery and were interviewed a week later about their recollections of the event, reporting their confidence for each recalled detail using a 1-to-7 scale. As illustrated in Figure 1, confidence was strongly related to accuracy, and high confidence was associated with high accuracy.

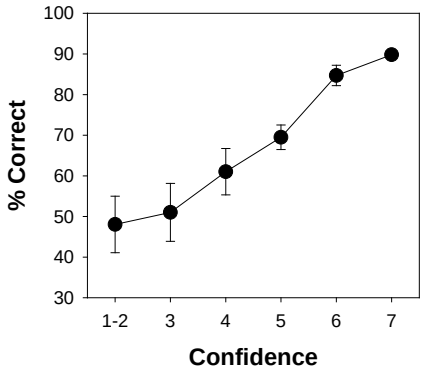

Figure 1. Accuracy (percent correct) as a function of confidence for the interview recall data reported by Roberts and Higham (2002). We estimated the overall number of correct details (considered relevant to an investigation) and incorrect details for each level of confidence from their Figure 1 and then computed the probability that a recalled detail was correct for each level of confidence. Adapted from Wixted et al. (2018). Error bars represent standard error.

In a study by Odinot, Wolters, & van Giezen (2013), participants watched a television show about an accident between a car and a motorcycle. One, three, or five weeks later, they completed a questionnaire consisting of 28 open-ended questions, and they rated their confidence in each recalled detail using 7-point scale. The results showed not only that confidence was

strongly indicative of accuracy, but high confidence implied high accuracy in all three conditions (Figure 2).

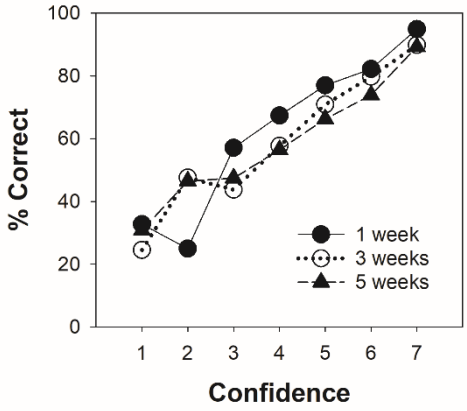

Figure 2. Accuracy (percent correct) as a function of confidence estimated from the interview recall data when recall took place 1, 3, or 5 weeks after encoding reported in Figure 1 of Odinot et al. (2013). Note that retention interval is an estimator variable (variables that affect memory but are outside the control of law enforcement), but its effect of the confidence-accuracy relationship was minor.

In recent study by Spearing and Wade (2021), participants watched a mock-crime video and rated confidence immediately following each freely recalled detail (Immediate) or after all information had first been recalled (Delayed). Once again, the relationship between confidence and accuracy was impressive, and high confidence implied high accuracy (Figure 3).[2]

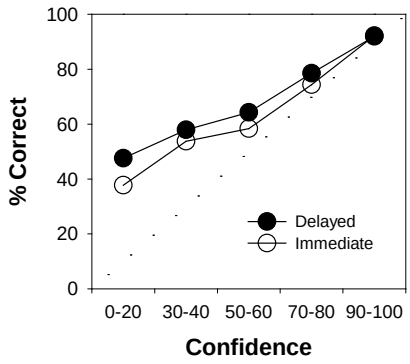

Figure 3. Accuracy (percent correct) as a function of confidence for the interview recall data reported by Spearing and Wade (2021). The data were estimated from their Figure 4 (Experiment 3). Experiment 3 was deemed by the authors to be the most forensically relevant of the three experiments they reported because it involved a free recall narrative report.

[2] Similar but more variable free recall results were reported in a smaller study by Brewer et al. (2018). In their Experiments 1 and 2, the accuracy of fine-grained details freely recalled with 100% confidence exceeded 90% correct, but in their Experiment 3 (involving only 14 participants), it was 78% correct.

Spearing and Wade (2021) also found that when the interview questions were misleading, high-confidence accuracy was only ~60% correct, confirming that improper testing procedures compromise the information value of eyewitness memory, just as improper DNA testing procedures compromise the information value of a DNA match. But when proper testing procedures are used, the relationship is usually impressive.

Do such findings generalize to the real world? No perfect study exists, but one is a reasonably close approximation. Odinot, Wolters, & van Koppen (2009) interviewed witnesses of an armed robbery that was recorded by multiple security cameras, and the recordings were used to verify the recalled information. The witnesses were interviewed three months after the crime, which is not ideal because considerable memory contamination can occur over that period (e.g., from conversations between witnesses, TV coverage, etc.). Yet Figure 4 shows that proportion correct increased from .61 to .85 as confidence increased from low to high. These numbers might be even more impressive had the interviews been conducted immediately after the crime, minimizing the opportunity for memory contamination. So far as we know, this the only interview study that has quantified confidence vs. accuracy in the real-world.

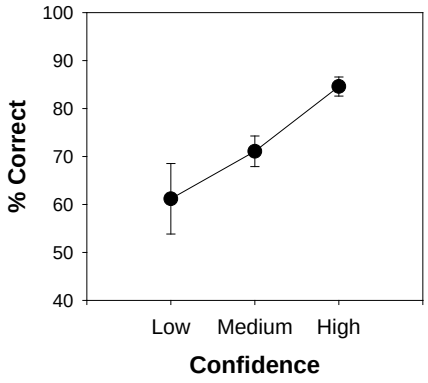

Figure 4. Observed relationship between percentage of recalled detailed that were correct and confidence. The data are from Odinot, Wolters, & van Koppen (2009). Adapted from Wixted et al. (2018). Error bars represent standard error.

Memory tested using a photo array (recognition of a face)

The same story applies to eyewitness memory tested by recognition using a lineup procedure (Wells et al., 2020). A proper procedure involves one suspect and five or more physically similar fillers, and the witness is cautioned that the perpetrator may or may not be among them. In addition, the witness's ID and immediate confidence statement must reflect a memory-based decision, not the influence of the lineup administrator, which is best assured using a double-blind procedure.

As is now well known, the field once incorrectly believed that eyewitness memory is unreliable even under "pristine" testing conditions. The analytical mistake was to measure the confidence-accuracy relationship using a correlation coefficient, which can be close to zero even when the relationship is nearly perfect (Juslin et al., 1996). Later studies used a calibration approach in which accuracy is plotted as a function of confidence (e.g., Brewer et al., 2002). Using that approach, the relationship was found to be stronger than previously believed (e.g., Brewer & Wells, 2006). Even so, the accuracy of IDs made with high confidence was not particularly impressive even under ideal testing conditions (e.g., 65% to 90% correct).

The dependent measure in a calibration plot is the probability that the identified suspect is guilty given that an ID was made to a suspect or a filler. However, suspect IDs are probative of guilt, whereas filler IDs are probative of innocence (Wells et al., 2015). Thus, those two decision outcomes are properly analyzed *separately* as a function of confidence, not mixed together. Unfortunately, the mistake of conflating them has made its way into the courtroom. For example, in an amicus brief, the American Psychological Association (APA) considered the calibration data reported by Sauer et al. (2008) and concluded 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 confident in the accuracy of their identifications…This confirms that many witnesses are over-confident in their identification decisions" (American Psychological Association, 2014, pp. 17–18).

However, the information of interest to judges and jurors is the probability that a *suspect* ID is correct as a function of confidence, which is the information provided by a confidence-accuracy characteristic (CAC) plot (Mickes, 2015). As shown in Figure 5, when the Sauer et al. (2008) data are analyzed this way, for IDs made with 90-100 confidence, accuracy was 96.8% correct (not 60% correct).

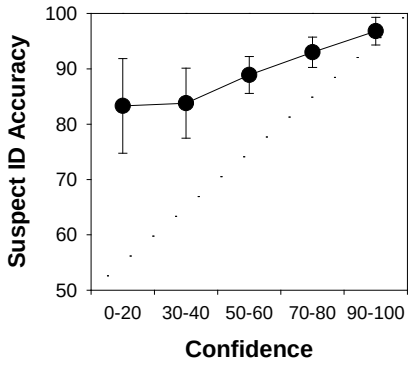

Figure 5. Suspect ID accuracy (percent correct) as a function of confidence for the data reported by Sauer et al. (2008). The data were collapsed across their Thief and Waiter conditions. Error bars represent standard error.

Wixted and Wells (2017) reviewed the larger literature, replotting all the previously collected correlation and calibration data in terms of CAC. As shown in Figure 6, when tested using a proper lineup procedure, confidence is predictive of accuracy, with high-confidence suspect IDs being very accurate. Much of the reanalyzed data had previously been used to argue that eyewitness identification accuracy is error-prone even for IDs made with high confidence. The fact that the *same data* show otherwise should be regarded as compelling evidence that what was previously thought to be true is in fact not true.

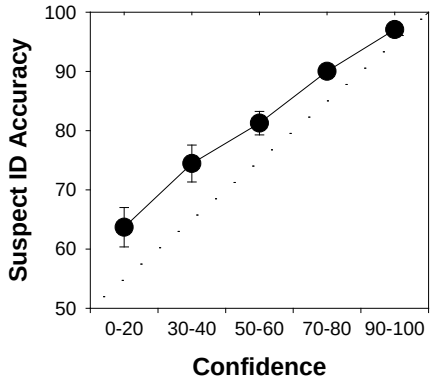

Figure 6. Suspect ID accuracy (percent correct) as a function of confidence averaged across 15 studies with comparable scaling on the confidence (x) axis (Wixted & Wells, 2017). Error bars represent standard error.

According to a recent study of actual eyewitnesses making IDs from properly administered fair lineups (Wixted et al., 2016), these results generalize to the real world (Figure 7).

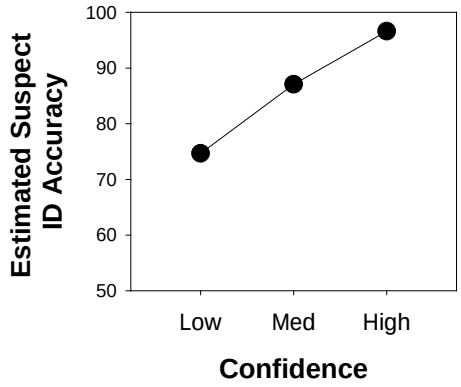

Figure 7. Estimated suspect ID accuracy (percent correct) as a function of confidence for the data from the Houston Police Department field study (Wixted et al., 2016). The crimes were robberies, the lineups were fair, the lineups were administered by an officer who was blind to the identity of the suspect, and the suspects in the lineups were not previously known to the witnesses.

The high accuracy of IDs made with high confidence is consistent with a key finding from the Hennepin County police department field study (Klobuchar et al., 2006), which found that, for fair, blindly administered lineups in which the suspect was a stranger to the eyewitness, 96% of “jump-out” IDs—that is, IDs indicative of absolute certainty—landed on the suspect (chance = 1/6 × 100% = 16.7%). No other real-world studies measuring eyewitness confidence

from properly administered lineups and have yet been conducted, but the findings from the studies conducted so far are in line with the lab findings.

Memory tested in the wild

Our message focuses strictly on science, according to which eyewitness memory is reliable when it is uncontaminated and properly tested, both in the lab and in the real world. How often the police properly test memory in the real world is an entirely separate matter, one that we have not addressed, either here or elsewhere. Undoubtedly, memory is tested properly in some police investigations and not in others. In any given criminal case, judges and jurors—not scientists who consider eyewitness memory to be presumptively unreliable regardless of individual case details—must determine how appropriate the testing procedure was and then judge the reliability of an eyewitness's memory accordingly. This means that, sometimes, the science-informed conclusion will presumably be that there is a high probability, though not a 1.0 probability, that an initial ID made with high confidence was accurate.

Words like those are not often written, and they have touched a nerve. For example, to Berkowitz et al. (2020), our message sounds like "confidence trumps all." But that is not our message, and it never has been (Wixted et al., 2021). Our message is not that eyewitness memory is so special that it is impervious to contamination and/or improper testing. Instead, our message is that eyewitness memory is *not* special. Like forensic DNA evidence, eyewitness memory is reliable when it is not contaminated and is properly tested. In our view, the data overwhelmingly support this claim.

It is entirely understandable that a field that has spent the better part of 50 years cautioning the legal system about the fallibility of eyewitness memory has not enthusiastically

embraced our message, nor should it. After all, scientists should be skeptical of any radical new claim, and they should do their best to poke holes in it (e.g., Sauer et al., 2019; Wade et al., 2018). Consider, for example, another surprising claim we have recently made about suboptimal estimator variables (e.g., short exposure duration, high stress, etc.): contrary to what was long believed to be to true, when the relevant data are reanalyzed in terms of CAC, high-confidence suspect IDs are surprisingly accurate whether the estimator variable conditions are favorable or unfavorable (Semmler et al., 2018). Unpersuaded, scientists have conducted new studies showing that when estimator variable conditions are sufficiently extreme such that overall performance is very poor, the accuracy of high-confidence suspect IDs is measurably reduced (Giacona et al., 2021; Lockamyeir et al., 2020).

As noted by Mickes and Wixted (in press), there are good theoretical reasons to expect this to be true. However, the new push-it-to-the-limit lab data may not be relevant in an applied sense. Based on a reanalysis of the real-world data reported by Wixted et al. (2016), Giacona et al. (2021, Fig. 4E) reported that, for actual eyewitnesses who expressed high confidence in an ID, the probability that it was a *suspect* ID was .86 under good viewing conditions and .77 under poor viewing conditions (again, chance = .167). According to the simplest estimation method used by Wixted et al. (2016, Equations 1 through 7), these values translate to high-confidence suspect ID accuracy estimates of 97.7% correct under good viewing conditions and 96.5% under poor viewing conditions.

These findings suggest that high-confidence initial IDs are highly reliable in the lab and in the real world, whether the estimator variable conditions are favorable or unfavorable, consistent with a great deal of lab-based research (Semmler et al., 2018). Keep in mind that for these real-world lineups, the estimator variables were often less-than-perfect in multiple respects.

For example, most involved the presence of a weapon (perhaps eliciting weapon focus and presumably eliciting high stress), and most involved cross-race IDs. Yet no matter how the data are analyzed, high confidence implies high accuracy.

Eyewitness identification researchers are understandably motivated to find conditions in the lab where this claim fails even on an initial, properly conducted test, but the mere fact that such conditions can be found does not automatically mean that they are relevant to the real world. Until it is shown that they are, it seems fair to suggest that, when tested properly, eyewitness memory may be far more reliable that many people believe. Moreover, by routinely testing memory improperly, the criminal justice system itself is responsible for the wrongful convictions ordinarily attributed to the fallibility of eyewitness memory. We suggest it is time to update textbooks and revise what we teach our students accordingly.

References

American Psychological Association. (2014). Commonwealth v. Gomes and Commonwealth v. Johnson. Retrieved fromhttp://www.apa.org/about/offices/ogc/amicus/gomes-johnson.

Berkowitz, S. R., Garrett, B. L., Fenn, K. M., & Loftus, E. F. (2020). Convicting with confidence? Why we should not over-rely on eyewitness confidence. *Memory.* Advance online publication. https://doi.org/10.1080/09658211.2020.1849308

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8*, 44–56.

Brewer, N., Vagadia, A. N., Hope, L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and Human Behavior, 42*, 458–471.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship 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.

Giacona, A. M., Lampinen, J. M., & Anastasi, J. S. (2021). Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law and human behavior, 45(3)*, 256–270.

Innocence Project (2021). Understand the causes: the causes of wrongful conviction. New York: Innocence Project. https://www.innocenceproject.org/eyewitness-identification-reform/. Accessed August 11, 2021.

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*(5), 1304–1316

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455(7210),* 227–231.

Kiani, R., Corthell, L. & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron, 84 (6),* 1329-1342.

Kiani, R. & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324,* 759–764.

Klobuchar, A., Steblay, N. K., & Caligiuri, H. L. (2006). Improving eyewitness identification: Hennepin County's blind sequential lineup pilot project. *Cardozo Public Law, Policy, and Ethics Journal, 2,* 381–414.

Lockamyeir, R. F., Carlson, C. A., Jones, A. R., Carlson, M. A., & Weatherford, D. R. (2020). The effect of viewing distance on empirical discriminability and the confidence-accuracy relationship for eyewitness identification. *Applied Cognitive Psychology, 34(5),* 1047–1060.

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*, 93-102.

Mickes, L. & Wixted, J. T. (in press). Eyewitness memory. In M. J. Kahana & A. D. Wagner (Eds.) *Oxford Handbook of Human Memory*. Oxford University Press.

Moran, R., Teodorescu, A. R. & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive Psychology, 78*, 99-147.

Odinot, G., Memon, A., La Rooy, D., & Millen, A. (2013). Are two interviews better than one? eyewitness memory across repeated cognitive interviews. *PloS one, 8*(10), e76305. https://doi.org/10.1371/journal.pone.0076305

Odinot, G., Wolters, G., & Lavender, T. (2009). Repeated partial eyewitness questioning causes confidence inflation but not retrieval-induced forgetting. Applied Cognitive Psychology, 23(1), 90–97.

Odinot, G., Wolters, G. & van Koppen, P. J. (2009). Eyewitness memory of a supermarket robbery: A case study of accuracy and confidence after 3 months. *Law and Human Behavior, 33*, 506-514.

Odinot, G., Wolters, G., & van Giezen, A. (2013). Accuracy, confidence and consistency in repeated recall of events. *Psychology, Crime & Law, 19*, 629-642.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117 (3),* 864-901.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review, 120*(3), 697–719.

Roberts, W. T. & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied, 8,* 33–43.

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

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

Semmler, C., Dunn, J., Mickes, L. & Wixted, J. T. (2018). The role of estimator variables in eyewitness identification. *Journal of Experimental Psychology: Applied, 24*, 400-415.

Smith, A. M., Wilford, M. M., Quigley-McBride, A., & Wells, G. L. (2019). Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior, 43(4)*, 358–368.

Spearing, E. R., & Wade, K. A. (2021). Providing eyewitness confidence judgements during versus after eyewitness interviews does not affect the confidence-accuracy relationship. *Journal of Applied Research in Memory and Cognition*. doi: https://doi.org/10.31234/osf.io/bm2ua

Steblay, N. K., Tix, R. W., & Benson, S. L. (2013). Double exposure: The effects of repeated identification lineups on eyewitness accuracy. *Applied Cognitive Psychology, 27*, 644–654

Wade, K. A., Nash, R. A., & Lindsay, D. S. (2018). Reasons to doubt the reliability of eyewitness memory: Commentary on Wixted, Mickes, and Fisher (2018). *Perspectives on Psychological Science, 13*, 339-342.

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*, 3-36.

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior*, *39*(2), 99–122.

Wixted, J. T., Mickes, L., Brewin, C. R., & Andrews, B. (2021). Doing Right by the Eyewitness Evidence: A Response to Berkowitz et al.. *Memory*. doi: 10.1080/09658211.2021.1940206

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D. & Roediger, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70*, 515-526.

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

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

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

Wixted, J. T., Wells, G. L., Loftus, E. F., & Garrett, B. L. (in press). Test a witness's memory for a suspect only once. *Psychological Science in the Public Interest*.