

# Proof Central

---

Please use this PDF proof to check the layout of your article. If you would like any changes to be made to the layout, you can leave instructions in the online proofing interface. First, return to the online proofing interface by clicking "Edit" at the top page, then insert a Comment in the relevant location. Making your changes directly in the online proofing interface is the quickest, easiest way to correct and submit your proof.

Please note that changes made to the article in the online proofing interface will be added to the article before publication, but are not reflected in this PDF proof.



Contents lists available at ScienceDirect

## Journal of Applied Research in Memory and Cognition

journal homepage: [www.elsevier.com/locate/jarmac](http://www.elsevier.com/locate/jarmac)

## Commentary

## Eyewitness Identification can be Studied in Social Contexts Online with Large Samples in Multi-lab Collaborations

D. Stephen Lindsay\* and Eric Y. Mah

University of Victoria

Keywords: Eyewitness; Identification; Police; Social context

Police procedures for collecting eyewitness identification evidence have been greatly influenced by scientific psychology (Wells, 2020). The history of psychological research in this domain is long, including publications in the late 19th and early 20th centuries by pioneers such as Binet, Munsterberg, Stern, and Whipple (Sporer, 2008). The impact of much of that research appears to have been modest until a few decades ago. Britain broke ground with the 1984 Police and Criminal Evidence Act (Bull, 2004). In 1999 the U.S. Department of Justice released *Eyewitness Evidence: A Guide for Law Enforcement* (National Institute of Justice, 1999). That guide was greatly influenced by Gary L. Wells and other psychologists on the Technical Working Group for Eyewitness Identification Evidence. As yet another example, in 2015 the Department of Justice Canada released its Report on the Prevention of Miscarriages of Justice, including many best-practice recommendations for collecting identification evidence based in large part on psychology research.

In response, police services in the US, Canada, and elsewhere dramatically changed procedures for collecting identification evidence. England and Wales have dedicated facilities for creating and presenting sequential video lineups via computer with guidance by a specially trained officer blind to which member of the lineup is the suspect (e.g., VIPER™ or Video Identification Parade Electronic Recording). Many police forces in the US, Canada, and other countries use sequential presentation of photospread lineups and other practices arising from psychology research intended to reduce mistaken identifications.

Most of these reforms seem to us to have been for the good. Grossly suggestive, unreliable procedures have largely been replaced by superior methods. But some of the recommenda-

tions that psychologists urged may have been hasty. Psychological scientists may, for example, have underestimated the value of eyewitnesses' confidence as an indicator of accuracy, at least under "pristine" conditions (Wixted et al., 2015; Wixted & Wells, 2017; cf. Read et al., 1998). As another example, it is no longer clear that sequential lineups robustly enhance witnesses' ability to discriminate between guilty and innocent suspects relative to simultaneous lineups (Kaesler et al., 2020; Wixted & Mickes, 2014).

On these and other issues our view is that the scientific jury should continue deliberating best practices for collecting identification evidence, not rush to judgment. For one thing, some published studies fall short of current methodological standards (i.e., had smallish samples, were not preregistered, used weak measures, etc.). For another, the gulf between the procedures typical of psychological research and the experiences of real-life witnesses and police is vast.

Kovera and Evelo (2021) made a strong case for the value of studying eyewitness identification in social contexts. They argued that in recent years an influx of cognitive psychologists into the community of researchers studying eyewitness identification led to increased use of signal detection measures of identification performance. Kovera and Evelo particularly emphasized the use of receiver operating characteristic (ROC) curves, which demand large numbers of subjects (especially if each subject is tested on only one or a few lineups). Improved understanding of statistical power may also have encouraged increases in sample size (e.g., Bakker et al., 2016). Increased demand for large samples coincided with the rise of internet-based crowd-sourcing platforms that make it easy to test thousands of subjects in standardized procedures devoid of live person-to-person interaction. Kovera and Evelo called for caution in generalizing from such studies to real-world situations in which motivated police officers work with witnesses. We agree.

\* Correspondence concerning this manuscript should be addressed to D. Stephen Lindsay, Department of Psychology, University of Victoria, Victoria, B.C. V8W 2Y2, Canada. Contact: [slindsay@uvic.ca](mailto:slindsay@uvic.ca).

## Playing at Science

In a thought-provoking paper, Syed (2021) argued that social psychology has been fettered and White-washed by an undue emphasis on experiments. Syed's primary thesis was that the focus on experiments has worked against ethnic and racial diversification of research in social psychology, but the argument (consonant with earlier critiques by the likes of Neisser, 1978, or Shweder, 1999) has more general implications for how we construe our task as psychological scientists.<sup>1</sup> "The desire to conduct experiments," Syed wrote "required that studies focus on micro contexts in the lab" (p. 3). He went on to note that in practice the experiments themselves, in addition to that narrow focus, suffered methodological/statistical weaknesses that (ironically) undercut the internal validity that experiments are supposed to deliver. As Syed put it, "This looks like playing at science" (p. 5).

The same charge of "playing at science" could be leveled at much of the published literature on eyewitness identification (including most efforts to date in our lab). Real witnesses see culprits in a great variety of contexts and are tested under a wide range of conditions. But no real-world witness has viewed a video of the crime and then minutes later performed a computer-administered lineup in exchange for bonus points in an introduction to psychology course. Granted, some studies do capture some aspects of real-world eyewitness situations (including some recently published work, such as Pike et al., 2019, and Rubínová et al., 2021). Also, sometimes principles observed in artificial lab studies turn out to generalize pretty well (Banaji & Crowder, 1989). But we agree with Kovera and Evelo that it is important for psychologists to study eyewitness suspect identification phenomena in socially embedded contexts that include motivated participant-investigators interacting with participant-witnesses.

As noted earlier, police in many jurisdictions use variants of the sequential lineup with photos. But we suspect that their practices vary from computer-based tests in many ways. Especially in smaller cities, it may be difficult to ensure that the administering officer is blind to the identity of the suspect, because they may be familiar with the suspect from other encounters and because photos of fillers are sometimes drawn from a pool of locals with prior histories with the police. Moreover, in the real world the test procedure may often have a conversational character. When the officer displays the first face in the deck, the witness may say "Yes" or "No," but probably they more often look at the photo, glance at the officer, return their gaze to the photo, and say something like "Hm. Well, now, I do remember he had curly hair. . .but still, I'm not sure. Can you show me the next one?" If by the end of the deck the witness hasn't made a positive identification, they may well say "Can I see the third and fourth ones again?," and the officer is likely to comply.

Our point here does not differ in substance from Kovera and Evelo's (2021) central thesis. If we want to speak with confi-

dence to best practices in real-world lineup identification tests, then we must increase the similarity between what we do in studies and what happens in the field. Kovera and Evelo argued that this can be done by putting the social context back into the laboratory. But they briefly touched on an alternate perspective that we think warrants serious consideration—the idea that "...it may be best to remove the social interaction from the procedure altogether by using computerized methods of lineup administration (p. 20) TYPESETTER PLEASE CHANGE TO ACTUAL PAGE NUMBER. In other words, taking the social context out of the field. The bevy of interviewer/administrator influences (mostly negative) described by Kovera and Evelo suggest that we might be better off enabling police to use non-biasing automated procedures for collecting identification evidence, akin to those used in online eyewitness experiments, rather than getting researchers to study social dynamics when non-blind lineup administrators interact freely with witnesses.<sup>2</sup>

One way to do that would be to get police to use computers to administer lineups much as researchers often do (Maclin et al., 2005). There is a lot to be said for that approach—consistent and controlled lineup viewing conditions, increased efficiency, decreased costs (Kemp et al., 2001), and evidence that live lineups are not superior to photo or video lineups (Rubínová et al., 2021). Kovera and Evelo (2021) reviewed numerous studies in which false identifications of innocent suspects were more common when administered by a participant-investigator who knew which photo was the suspect than when administration was double blind, which is perhaps most easily implemented via automation (see also Kovera & Evelo, 2020; Smalarz & Wells, 2015). So, we see merit in studying standardized/computerized lineup methods as a potential way to mitigate administrator effects and address the documented variability in police lineup procedures across jurisdictions (e.g., Wogalter et al., 2004).

That said, we acknowledge that there may be barriers to widespread implementation of automated lineup procedures, especially for police forces in smaller communities (as noted by Haw & Fisher, 2004). Moreover, we have the impression that some experienced police officers are resistant to the use of double-blind procedures or automated procedures. Further, we suspect that this resistance arises in part from police officers' belief that they can help witnesses make accurate identifications of guilty suspects without (much) increasing the risk of false identifications of innocent suspects.<sup>3</sup>

It would be interesting to know more about real-life identification practices in various parts of the world. The Police Executive Research Forum (2013) reported a major survey of a random stratified sample of more than 600 law enforcement agencies in the United States regarding their eyewitness identification practices. Most of the surveys were returned in 2011, and it may be that practices have changed considerably since then, especially in larger forces. Moreover, these data (valuable

<sup>2</sup> Of course, there are important differences between performing a computer-controlled identification procedure for the culprit of a real crime versus performing one as part of an online experiment.

<sup>3</sup> How likely that is depends in large part on how rarely police run lineups for innocent suspects (Cohen et al., 2020; Malpass, 2006).

<sup>1</sup> Thanks to Qi Wang for suggesting we cite Shweder in this context, leading us to read this mind-expanding paper. We don't claim to understand, let alone endorse, all of Shweder's arguments, but they are rich food for thought.

as they are) concern reported policy rather than police officers' beliefs and actual practices. Some studies that have examined beliefs have found that officers' beliefs can differ substantially from those of eyewitness research experts (Benton et al., 2006; Bertrand et al., 2018; Huang & Shi, 2020; Karagiorgakis, 2010; Tupper et al., 2019; Wise et al., 2011; Wogalter et al., 2004). We think it would be worthwhile to learn more about police officers' current self-reported beliefs and their in-the-trenches practices. And if there is good evidence that many police have mistaken beliefs, to learn more about how to persuade them to update and correct.

#### Pros and Cons of SDT Measures

Kovera and Evelo (2021) rightly observed that ROCs require large samples. That is a problem. The only worse problem is using measures that don't tell us what we want to know. Without question psychologists can glean valuable insights into eyewitness identification from other sorts of measures. But when it comes to determining which procedures yield better versus worse identification performance, at present the good options all require fairly large numbers of observations. It seems undesirable to rely on hit rates, false alarm rates, or diagnosticity ratios, because each of these measures can be misleading on its own (Gronlund et al., 2014).

Lampinen (2016) argued that ROC curves do not measure theoretical discriminability or accurately index the utility of lineup procedures. Wixted et al. (2017) countered that ROC curves provide a useful atheoretical measure of empirical discriminability, and when analyzed correctly provide information about the utility of lineup procedures. Simpler measures such as  $d'$  often yield conclusions similar to those gleaned from ROC analysis (and more accurate than diagnosticity ratios), but sometimes  $d'$  and ROC measures diverge and atheoretic ROCs arguably are the better index (Mickes et al., 2014). That said, we agree with Lampinen (2016) that lineup research benefits from multiple theories, measures, and analyses, and that ROC curves should not be our only measure of lineup outcomes. There may be reason to prefer  $d'$  if it requires smaller samples (but we are not aware of research directly comparing the power of  $d'$  to ROC curves). So, even if ROC measures are not the be-all-end-all, we think it premature to do away with the SDT framework.

Efforts are being made to increase the quality and ecological validity of current SDT-based measures. For instance, Smith and Neal (in press) argued that while discriminability is of primary interest to scientists, reliability (the trustworthiness of evidence) is the currency of the criminal justice system. They show that estimates of reliability can be extracted from SDT-based models. More generally, ROC curves are a relatively new measure in the eyewitness literature, and there are ongoing concerted efforts to improve their reliability and validity (e.g., Smith et al., 2019, 2020). Additionally, increasingly sophisticated SDT-based computational models of lineup outcomes show promise for addressing shortcomings of extant SDT-based measures (e.g., Colloff & Wixted, 2020; Lee & Penrod, 2019). Wixted et al. (2016) and Cohen et al. (2020) found that such models could be used to estimate quantities

highly relevant to real-world decision-making (e.g., the base rate of guilty suspects in lineups; the probability that a suspect is guilty given a witness identification).

If each subject views only one culprit and is tested on only one line-up (which seems desirable on grounds of ecological validity) then quite large sample sizes are required to obtain high statistical power to detect modest (yet practically important) effect sizes. That holds even if you just want to compare the rate of mistaken identifications of innocent suspects in two conditions (e.g., according to  $G^*$  Power 3 (Faul et al., 2007), having 80% power to detect a doubling in false IDs from 10% to 20% requires 214 subjects per group).

Researchers have proposed alternative lineup procedures that yield multiple data points per lineup, such as the rate-em-all confidence procedure of Sauer et al. (2008) and Brewer et al. (2020) and the rank-em-all procedure proposed by Carlson et al. (2019). Possibly these measures will prove to be more sensitive and reliable and less ambiguous than diagnosticity ratios and thereby reduce need for very large sample sizes.

As argued by Baldassari, Kantner, and Lindsay (2019), it may be useful to combine identification judgments with other measures predictive of accuracy, such as speed of responding, self-reported confidence, witnessing conditions, delay, functional size of the lineup, and individual differences in face processing skill and in proclivity to choose. Perhaps eventually psychologists will develop formal models that enable police to more accurately update estimates of the strength of evidence of a suspect's guilt on the basis of a lineup response in the context of multiple predictors. In principle, social-context variables could be included in formal models of eyewitness behaviour.

SDT-based measures that do not require very large numbers of observations may be in the offing, but they are not here yet. All told, we think that it may be premature to do away with SDT-based measures of eyewitness memory. But we must find ways to situate such measures (with the requisite large samples) in the social contexts that Kovera and Evelo described.

#### Ways to Conduct Large-N Studies with Rich Social Contexts

Kovera and Evelo (2021) wrote that it is all-but-impossible to conduct large-N studies that explore how social interactions influence witnesses' decisions on lineups. Certainly, it is easier to collect large samples in highly constrained, artificial, mass-testing procedures (whether paper and pencil or online). But one positive outcomes of the COVID-19 pandemic has been the proliferation of internet platforms that enable face-to-face interpersonal interactions in real time. This technology lends itself to socially dynamic studies of eyewitness memory. We developed a procedure in which pairs of participants interact with one another and with an experimenter online. After introductions and informed consent via Zoom with cameras and mics on, the participants watch a video in which a suspicious character snoops around a building and steals a few items. Unbeknownst to them, subjects are shown subtly different versions of the video. The two subjects then collaborate to answer questions about critical details in the video. Some questions concern details for which each subject had been shown a differ-



ent answer. Later, they are tested individually. Data collection is ongoing, but behaviour in this setting appears to be similar to behaviour in prior studies conducted with pairs of subjects interacting face to face in real life (e.g., Gabbert et al., 2003; Ito et al., 2019).

Admittedly, using an online platform such as Zoom only slightly mitigates the difficulty of conducting large-N studies of live, person-to-person social interactions. It is a time-consuming, labour-intensive process to collect and code extended interactions between pairs of people. It might not be feasible for a researcher to test hundreds of participant-officers, each interacting with a participant-witness. Still, Zoom-based experiments preserve one advantage of crowdsourced online data collection: samples that are more diverse and representative than typical university research pools in the lab (Peer et al., 2017). Whether data collection occurs online or in the lab, injecting social context into experimental manipulations is only a part of the solution—we must also investigate these manipulations in the populations in which the social contexts occur in the real world, and that is not easy.

Happily, another recent methodological innovation may come to the rescue. It turns out that it is possible for researchers to collaborate. Each of multiple teams can develop a shared plan for a research project and then collect and code data that are combined for analysis. Such studies can be coordinated through the Psychological Science Accelerator, “a globally distributed network of psychological science laboratories (currently over 500), with over 1400 members representing 71 countries on all six populated continents, that coordinates data collection for democratically selected studies” (Moshontz et al., 2018; <https://psysciacc.org/>). Or investigators can informally reach out to fellow researchers to develop a consortium of collaborators for a particular project (for inspiring examples, see Byers-Heinlein et al., 2020; Vohs et al., in press; for information about a related approach known as registered replication reports, see Simons et al., 2014).

The call here is for a cultural shift from Lone Ranger researchers to distributed teams of scientists working in coordination on shared projects. Pooling resources in this way makes it more feasible to test and code large samples of subjects in time-consuming procedures (such as face-to-face dynamic interactions, whether online or in real life). Multiple-team projects may also lead to improved methods, as multiple researchers collaborate on planning a study. Fewer studies get done than if each lab conducted smaller independent studies, but the hope is that the collaborative projects will be more rigorous, more transparent, and more statistically powerful. Multi-lab projects might also add to the generality of findings, especially in the case of international collaborations or projects that leave some methodological details to be determined by individual labs. See Uhlmann et al. (2019) for more detailed discussion of these sorts of approaches (and for related efforts in the domain of longitudinal studies of cognitive aging, see <http://www.ialsa.org/>). We are not suggesting that all research should use this multi-lab approach, but we believe it worth considering as a way of conducting large-N studies in which sub-

ject-witnesses and subject-investigators have dynamic social interactions.

Another way to increase the ecological validity of eyewitness studies is the involvement of real officers, either as participants or advisors. In our lab, we have collaborated with our local PD on the creation of an Electronic Self-Administered Cognitive Interview that includes interview clips recorded by a uniformed officer (for other recent examples of fruitful collaborations between researchers and officers in the literature, see Sharps et al., 2009, and Vredeveltdt et al., 2015). In addition to increasing the ecological validity of experiments, such collaborations bridge the gap between researchers and practitioners. Some years ago, Mario Baldassari and Steve Lindsay launched a project in which experienced police officers interviewed student-witnesses regarding a video taped simulated crime, and later administered a lineup we provided. Each officer worked with four student-witnesses, each of whom had seen a different sort of crime video. Police knew which member of the lineup was the suspect, and in two of the lineups the suspect was in fact the culprit from the video whereas in the remaining two cases the lineup included an innocent suspect. Compared to student-investigators, police were vastly better at interviewing witnesses. But it was not clear if they were any better at differentiating between accurate and inaccurate lineup judgments, largely because after months of trying we abandoned the project having recruited only 15 police officers. With the advent of distributed lab networks, larger-sample studies with staged crimes and officer involvement may be more feasible than ever.

### Conclusion

In sum, we agree with Kovera and Evelo (2021) that the field must do more to situate research in the social contexts so central to real-world eyewitness experiences. As they suggested, one way to do this is to shift our focus from large-N, online, “pristine,” artificial, SDT-based experiments to applied studies that measure real interpersonal behavior. However, we argue that (a) online experiments can incorporate social contexts and manipulations, (b) that there is value in applied research investigating “impersonal” lineup procedures (i.e., taking the social context out of the real world rather than putting it into experiments), (c) given the proven worth of SDT-based measures and ongoing efforts to improve them, the best way forward may be to incorporate social context variables into the SDT framework (vs. doing away with it entirely), and finally and perhaps most excitingly (d) that distributed lab networks offer a promising new way to conduct large-N applied studies in socially rich contexts.

### Author Note

We thank Ryan J. Fitzgerald and Ira E. Hyman, Jr., for helpful comments on an earlier draft of this manuscript. Remaining shortcomings are our own.

## Author contributions

Each author independently generated ideas regarding points to be made, then shared those notes. Lindsay then drafted an initial version, which Mah revised.

## Conflict of interest statement

The authors declare that they have no conflict of interest.

## Uncited references

Mansour et al. (2017), Rubínová et al. (2020), Seale-Carlisle et al. (2019), Wise et al. (2009).

## References

- Bakker, M., Hartgerink, C. H. J., Wicherts, J. M., & van der Maas, H. L. J. (2016). Researchers' intuitions about power in psychological research. *Psychological Science* 27, 1069–1077. <https://doi.org/10.1177/0956797616647519>.
- Baldassari, M. J., Kantner, J., & Lindsay, D. S. (2019). The importance of decision bias for predicting eyewitness lineup choices: Toward a lineup skills test. *Cognitive Research: Principles and Implications*, 4. <https://rdocu.be/cnqY5>
- Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist* 44, 1185–1193. <https://doi.org/10.1037/0003-066X.44.9.1185>.
- Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, III, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology* 20, 115–129. <https://doi.org/10.1002/acp.1171>.
- Bertrand, M., Lindsay, R. C. L., Mansour, J. K., Beaudry, J. L., Kalmet, N., & Melson, E. J. (2018). Examining how lineup practices of Canadian and U.S. police officers adhere to their national best practice recommendations. *Manitoba Law Review* 41, 1–47.
- Brewer, N., Weber, N., & Guerin, N. (2020). Police lineups of the future? *American Psychologist* 75, 76–91. <https://doi.org/10.1037/amp0000465>.
- Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., Waddell, C. P. G., Zettersten, M., & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne* 61, 349–363. <https://doi.org/10.1037/cap0000216>.
- Carlson, C. A., Jones, A. R., Goodsell, C. A., Carlson, M. A., Weatherford, D. R., Whittington, J. E., & Lockamyeir, R. F. (2019). A method for increasing empirical discriminability and eliminating top-row preference in photo arrays. *Applied Cognitive Psychology* 33, 1–12. <https://doi.org/10.1002/acp.3551>.
- Cohen, A. L., Starns, J. J., Rotello, C. M., & Cataldo, A. M. (2020). Estimating the proportion of guilty suspects and posterior probability of guilt in lineups using signal-detection models. *Cognitive Research: Principles and Implications* 5. <https://doi.org/10.1186/s41235-020-00219>.
- Department of Justice Canada (2015). Report on the prevention of miscarriages of justice. <https://www.justice.gc.ca/eng/tp-pr/cj-jp/ccr-rc/pmj-pej/pmj-pej.pdf>.
- Faul, F., Erdfelder, E., Lange, A.-G., & Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39(2), 175–191. <https://doi.org/10.3758/bf03193146>.
- Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology* 17, 533–543.
- Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science* 23, 3–10. <https://doi.org/10.1177/0963721413498891>.
- Haw, R. M., & Fisher, R. P. (2004). Effects of administrator-witness contact on eyewitness identification accuracy. *Journal of Applied Psychology* 89, 1106–1112. <https://doi.org/10.1037/0021-9010.89.6.1106>.
- Huang, C.-Y., & Shih, C.-H. (2020). The good, the bad and the ugly of eyewitness identification practice in police officers – a self-report survey study. *Psychology, Crime & Law* 26, 1006–1026. <https://doi.org/10.1080/1068316X.2020.1744602>.
- Ito, H., Barzykowski, K., Grzesik, M., Gülgöz, S., Gürdere, C., Janssen, S. M. J., Khor, J., Rowthorn, H., Wade, K. A., Luna, K., Albuquerque, P. B., Kumar, D., Singh, A. D., Ceconello, W. W., Cadavid, S., Laird, N. C., Baldassari, M. J., Lindsay, D. S., & Mori, K. (2019). Eyewitness memory distortion following co-witness discussion: A replication of Garry, French, Kinzett, and Mori (2008) in ten countries. *Journal of Applied Research in Memory and Cognition* 8, 68–77.
- Karagiorgakis, A. (2010). Police Officer Beliefs about Factors that Influence Eyewitness Memory. Unpublished doctoral dissertation.
- Kaesler, M., Dunn, J. C., Ransom, K., & Semmler, C. (2020). Do sequential lineups impair underlying discriminability? *Article 35 Cognitive Research: Principles and Implications* 5. <https://doi.org/10.1186/s41235-020-00234-5>.
- Kemp, R. I., Pike, G. E., & Brace, N. A. (2001). Video-based identification procedures: Combining best practice and practical requirements when designing identification systems. *Psychology, Public Policy, and Law* 7, 802–807. <https://doi.org/10.1037/1076-8971.7.4.802>.
- Kovera, M. B., & Evelo, A. J. (2020). Improving eyewitness-identification evidence through double-blind lineup administration. *Current Directions in Psychological Science* 29, 563–568.
- Kovera, M. B., & Evelo, A. J. (2021). Eyewitness identification in its social context. *Journal of Applied Research in Memory and Cognition*. <https://doi.org/10.1016/j.jarmac.2021.04.003>.
- Lampinen, J. M. (2016). ROC analyses in eyewitness identification research. *Journal of Applied Research in Memory and Cognition* 5, 21–33. <https://doi.org/10.1016/j.jarmac.2015.08.006>.
- Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior* 43, 436–454. <https://doi.org/10.1037/lhb0000343.supp>.
- MacLin, O. H., Meissner, C. A., & Zimmerman, L. A. (2005). PC\_Eyewitness: A computerized framework for the administration and practical application of research in eyewitness psychology. *Behavior Research Methods* 37, 324–334.
- Malpass, R. S. (2006). A policy evaluation of simultaneous and sequential lineups. *Psychology, Public Policy, and Law* 12, 394–418.
- Mansour, J. K., Beaudry, J. L., & Lindsay, R. C. L. (2017). Are multiple-trial experiments appropriate for eyewitness identification

- studies? Accuracy, choosing, and confidence across trials. *Behavior Research Methods* 49, 2235–2254. <https://doi.org/10.3758/s13428-017-0855-0>.
- Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute  $d'$ , not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition* 3, 58–62. <https://doi.org/10.1016/j.jarmac.2014.04.007>.
- Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J., Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., & ... Chartier, C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science* 1, 501–515. <https://doi.org/10.1177/2515245918797607>.
- National Institute of Justice, (1999). *Eyewitness evidence: A guide for law enforcement*. Office of Justice Programs, Washington, DC.
- Neisser, U. (1978). Memory: What are the important questions? In *Practical aspects of memory* (eds. M. M. Gruneberg, P. E. Morris and R. N. Sykes). Academic Press, London, pp. 3–24.
- Pike, G. E., Brace, N. A., Turner, J., & Vredeveldt, A. (2019). The effect of facial composite construction on eyewitness identification accuracy in an ecologically valid paradigm. *Criminal Justice and Behavior* 46(2), 319–336. <https://doi.org/10.1177/0093854818811376>.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>.
- Police Executive Research Forum. (2013). A national survey of eyewitness identification procedures in law enforcement agencies. <https://www.ncjrs.gov/pdffiles1/nij/grants/242617.pdf>.
- Read, J. D., Lindsay, D. S., & Nichols, T. (1998). The relation between confidence and accuracy in eyewitness identification studies: Is the conclusion changing? In *Eyewitness memory: Theoretical and applied perspectives* (eds. C. P. Thomson, D. Bruce, J. D. Read, D. Hermann, D. Payne and M. P. Toglia). Lawrence Erlbaum, Mahwah, NJ, pp. 107–130.
- Rubínová, E., Fitzgerald, R. J., Juncu, S., Ribbers, E., Hope, L., & Sauer, J. D. (2020). Live presentation for eyewitness identification is not superior to photo or video presentation. *Journal of Applied Research in Memory and Cognition*. <https://doi.org/10.1016/j.jarmac.2020.08.009>.
- Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General* 137, 528–547. <https://doi.org/10.1037/a0012712>.
- 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, 420–428. <https://doi.org/10.1016/j.jarmac.2019.09.003>.
- Sharps, M. J., Janigian, J., Hess, A. B., & Hayward, B. (2009). Eyewitness memory in context: Toward a taxonomy of eyewitness error. *Journal of Police and Criminal Psychology* 24, 36–44. <https://doi.org/10.1007/s11896-008-9029-4>.
- Shweder, R. A. (1999). Why cultural psychology? *Ethos* 27, 62–73. <https://doi.org/10.1525/eth.1999.27.1.62>.
- Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at perspectives on psychological science. *Perspectives on Psychological Science* 9, 552–555. <https://doi.org/10.1177/1745691614543974>.
- Smalarz, L., & Wells, G. L. (2015). Contamination of eyewitness self-reports and the mistaken-identification problem. *Current Directions in Psychological Science* 24, 120–124. <https://doi.org/10.1177/0963721414554394>.
- Smith, A. M., Lampinen, J. M., Wells, G. L., Smalarz, L., & Mackovichova, S. (2019). Deviation from perfect performance measures the diagnostic utility of eyewitness lineups but partial area under the ROC curve does not. *Journal of Applied Research in Memory and Cognition* 8, 50–59. <https://doi.org/10.1016/j.jarmac.2018.09.003>.
- Smith, A. M., Yang, Y., & Wells, G. L. (2020). Distinguishing between investigator discriminability and eyewitness discriminability: A method for creating full receiver operating characteristic curves of lineup identification performance. *Perspectives on Psychological Science* 15, 589–607. <https://doi.org/10.1177/1745691620902426>.
- Smith, A.M., & Neal, T.M.S. (In press). The distinction between discriminability and reliability in forensic science. *Science & Justice*. Doi: 10.1016/j.scijus.2021.04.002
- Sporer, S. L. (2008). Lessons from the origins of eyewitness testimony research in Europe. *Applied Cognitive Psychology* 22(6), 737–757. <https://doi.org/10.1002/acp.1479>.
- Syed, M. (2021). It's 2 x 2 designs all the way down: Social psychology's over-reliance on experiments needlessly restricts diversity in the field. Doi: 10.31234/osf.io/u89e2.
- Tupper, N., Sauerland, M., Sauer, J. D., & Hope, L. (2019). Eyewitness identification procedures for multiple perpetrator crimes: A survey of police in Sweden, Belgium, and the Netherlands. *Psychology, Crime & Law* 25(10), 992–1007. <https://doi.org/10.1080/1068316X.2019.1611828>.
- Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., McCarthy, R. J., Riegelman, A., Silberzahn, R., & Nosek, B. A. (2019). Scientific utopia III: Crowdsourcing science. *Perspectives on Psychological Science* 14 (5), 711–733. <https://doi.org/10.1177/1745691619850561>.
- Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., Alquist, J. L., Baker, M. D., Brizi, A., Buniy, A., Butschek, G. J., Campbell, C., Capaldi, J., Cau, C., Chambers, H., Chatzisarantis, N. L. D., Christensen, W. J., Clay, S. L., Curtis, J., ..., & Albarracín, D. (In press). A multi-site preregistered paradigmatic test of the ego depletion effect. 10.31234/osf.io/e497p.
- Vredeveldt, A., Tredoux, C. G., Nortje, A., Kempen, K., Puljević, C., & Labuschagne, G. (2015). A field evaluation of the eye-closure interview with witnesses of serious crimes. *Law and Human Behavior* 39, 189–197. <https://doi.org/10.1037/lhb0000113>.
- Wells, G. L. (2020). Psychological science on eyewitness identification and its impact on police practices and policies. *American Psychologist* 75, 1316–1329. <https://doi.org/10.1037/amp0000749>.
- Wise, R. A., Safer, M. A., & Maro, C. M. (2009). What U.S. law enforcement officers know and believe about eyewitness factors, eyewitness interviews, and identification procedures. *Applied Cognitive Psychology* 25, 488–500. <https://doi.org/10.1002/acp.1717>.

- 637 Wixted, J. T., & Mickes, L. (2014). A signal-detection-based  
638 diagnostic-feature-detection model of eyewitness identification.  
639 Psychological Review 121, 262–276. [https://doi.org/10.1037/  
640 a0035940](https://doi.org/10.1037/a0035940).
- 641 Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger,  
642 H. L. (2015). Initial eyewitness confidence reliably predicts  
643 eyewitness identification accuracy. American Psychologist 70,  
644 515–526. <https://doi.org/10.1037/a0039510>.
- 645 Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W.  
646 (2016). Estimating the reliability of eyewitness identifications from  
647 police lineups. Proceedings of the National Academy of Sciences  
648 113, 304–309.
- 649 Wixted, J. T., Mickes, L., Wetmore, S. A., Gronlund, S. D., &  
650 Neuschatz, J. S. (2017). ROC analysis in theory and practice.  
Journal of Applied Research in Memory and Cognition 6, 343– 651  
351. <https://doi.org/10.1016/j.jarmac.2016.12.002>. 652
- Wixted, J. T., & Wells, G. L. (2017). The relationship between 653  
eyewitness confidence and identification accuracy: A new synthe- 654  
sis. Psychological Science in the Public Interest 18, 10–65. [https:// 655  
doi.org/10.1177/1529100616686966](https://doi.org/10.1177/1529100616686966). 656
- Wogalter, M. S., Malpass, R. S., & Mcquiston, D. E. (2004). A 657  
national survey of US police on preparation and conduct of 658  
identification lineups. Psychology, Crime & Law 10, 69–82. 659  
<https://doi.org/10.1080/10683160410001641873>. 660

Received July 2, 2021 661  
accepted July 2, 2021 662  
Available Online: xxxx 663  
664