



Short essay

## ROC analysis of lineups does not measure underlying discriminability and has limited value



Gary L. Wells<sup>a,\*</sup>, Laura Smalarz<sup>b</sup>, Andrew M. Smith<sup>c</sup>

<sup>a</sup> Iowa State University, Psychology Department, West 112 Lago Hall, Ames, IA 50011, United States

<sup>b</sup> Williams College, Williamstown, MA, United States

<sup>c</sup> Queen's University, Kingston, ON, Canada

### ARTICLE INFO

#### Article history:

Received 2 July 2015

Accepted 30 August 2015

Available online 3 September 2015

#### Keywords:

Eyewitness identification

Lineups

Showups

Receiver Operating Characteristic (ROC) analysis

Bayesian analysis

Lineup fillers

Lineup discriminability

### ABSTRACT

Some researchers have been arguing that eyewitness identification data from lineups should be analyzed using Receiver Operating Characteristic (ROC) analysis because it purportedly measures underlying discriminability. But ROC analysis, which was designed for  $2 \times 2$  tasks, does not fit the  $3 \times 2$  structure of lineups. Accordingly, ROC proponents force lineup data into a  $2 \times 2$  structure by treating false-positive identifications of lineup fillers as though they were rejections. Using data from lineups versus showups, we illustrate how this approach misfires as a measure of underlying discriminability. Moreover, treating false-positive identifications of fillers as if they were rejections hides one of the most important phenomena in eyewitness lineups, namely filler siphoning. Filler siphoning reduces the risk of mistaken identification by drawing false-positive identifications away from the innocent suspect and onto lineup fillers. We show that ROC analysis confuses filler siphoning with an improvement in underlying discriminability, thereby fostering misleading theoretical conclusions about how lineups work.

© 2015 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.

Eyewitness identification experiments in psychology flagged eyewitness identification as a potential source of convictions of the innocent long before the advent of forensic DNA testing. And, although DNA-based exonerations of the innocent show that approximately 75% involved mistaken identifications (Innocence Project, 2015) and field studies show that actual witnesses identify an innocent person approximately 1/3rd of the times that they identify someone (Wells, Steblay, & Dysart, 2015a), the controlled laboratory experiment remains the primary tool for theoretical and applied advances in eyewitness identification. Hence, it is essential that eyewitness researchers openly discuss various issues that arise regarding laboratory methods used to reach conclusions about eyewitness identification.

One such issue that has arisen in the last few years is the use of Receiver Operating Characteristic (ROC) curves for analyzing eyewitness performance in lineups. ROC analysis (based on Signal Detection Theory) and the Signal Detection parameter  $d'$  are generally considered to be largely interchangeable measures of underlying “discriminability” (see Mickes, Moreland, Clark, & Wixted, 2014). Discriminability refers to the ability of a perceiver to

distinguish between a signal (the true target) in noise versus noise alone (a lure). Stated another way, “Discriminability is represented by the degree of overlap between the target and lure distributions” (Wixted & Mickes, 2014, p. 264).

ROC proponents have been dismissive of any other way of examining lineup data: “It is not yet clear which lineup procedure will prove to be generally superior, but it is clear that ROC analysis is the only way to make that determination” (Gronlund, Wixted, & Mickes, 2014, p. 3). The only way? This strong claim is based on the belief among ROC proponents that ROC analyses of lineups are measures of underlying discriminability: “The lineup procedure that yields the highest empirical ROC is the one that theoretically best facilitates the discrimination between innocent and guilty suspects by reducing the overlap between the corresponding memory strength distributions” (Wixted & Mickes, 2014, p. 266). More recently, Mickes and Wixted said “discriminability refers to the ability to distinguish between the face that was seen in the video from faces that were not (including the face of an innocent suspect)” (Mickes & Wixted, 2015, p. 402).

In this article we show why ROC analysis of lineups is not a measure of underlying discriminability and how the application of ROC analysis to lineup data has led to inaccurate conclusions about underlying discriminability. Especially problematic is the way that ROC analysis obscures the ability to observe one of the

\* Corresponding author. Tel.: +1 515 294 6033.

E-mail address: [gjwells@iastate.edu](mailto:gjwells@iastate.edu) (G.L. Wells).

most fundamental phenomena in lineups, namely filler siphoning, which is essential to understanding how lineups actually work.

### 1. A recent example of how ROC analysis misfires when applied to lineups

It is important to note from the outset that we believe that ROC analysis is an appropriate estimate of discriminability when used on the problem structure for which it was developed, which is the classic  $2$  (signal + noise versus noise)  $\times 2$  (positive versus negative response) structure. But, as we will show, when applied to lineups (which have a  $3 \times 2$  structure) ROC analysis does not measure discriminability. Lineups have a natural  $3 \times 2$  structure because lineups contain a guilty suspect (culprit present) or innocent suspect (culprit absent) along with known-innocent individuals, or fillers. Hence, witnesses can make one of three responses from a lineup: a positive identification of the suspect, a positive identification of a filler, or a negative (rejection) response to the lineup. In developing the ROC approach to lineups, [Mickes, Flowe, and Wixted \(2012\)](#) faced a challenge, namely how to deal with filler identifications, which do not fit into the classic  $2 \times 2$  structure required for ROC analysis. Their solution was to force the  $3 \times 2$  structure of lineups into a  $2 \times 2$  structure by treating all mistaken filler identifications as rejections.

In a recent example of how this approach misfires as a measure of discriminability, we present data from [Wetmore et al. \(2015\)](#), who used ROC analysis to compare lineups to showups. Showups are identification procedures that present either a single guilty suspect (culprit present) or a single innocent suspect (culprit absent) and the witness makes either a positive identification or a negative (rejection) response. Hence, showups naturally fit the  $2 \times 2$  structure for which ROC analysis was developed. Following in the steps of [Mickes et al. \(2012\)](#), however, [Wetmore et al. \(2015\)](#) collapsed filler identifications from the lineup into the rejection category for purposes of conducting an ROC analysis. Consider the data from Wetmore et al. in Panel A of [Table 1](#). Notice that when the culprit was absent from the lineup, witnesses made a false-positive error 64.4% of the time (10.2% false positive identifications of the innocent suspect plus 54.2% false positive identifications of fillers). But, the ROC approach of collapsing filler identifications into the rejection category so as to create a  $2 \times 2$  representation yields the representation shown in Panel B. Panel B would have us believe that only 10.2% of witnesses made a false positive identification from the lineup when the culprit was absent. Panel B would therefore suggest that discriminability from the lineup was quite high because hits were high (68.3% of witnesses made a positive identification of the culprit when the culprit was present) and “correct rejections” were high (89.8% of witnesses managed to reject the lineup when the culprit was absent).

By conducting an ROC analysis using this misleading  $2 \times 2$  representation (Panel B in [Table 1](#)) of the lineup data, [Wetmore et al. \(2015\)](#) concluded that discriminability was better for the lineup than for the showup. After all, hits were roughly equivalent for the two procedures (68.3% for lineups vs. 62.1% for showups, a non-significant difference) and the  $2 \times 2$  representation of the lineup data made it appear that correct rejections were considerably higher for the lineup (89.8%) than for the showup (58.0%). Except that is not what happened. As is clear from Panel A, witnesses correctly rejected the lineup only 35.6% of the time when the culprit was absent, which is a *lower* correct-rejection rate than for the showup (58.0% correct rejections). Instead of improving witnesses' abilities to make a rejection response when the culprit was not there, the lineup simply led some of the witnesses who would have mistakenly identified the innocent suspect from the showup to mistakenly identify a filler from the lineup. That is not

better discriminability between signal and noise. That is simply shifting a mistaken identification from one person to another person. When the purpose is to estimate discriminability, it does not matter whether the identified person is an innocent suspect or an innocent filler; both are false-positive errors.

But by treating filler identifications as if they were rejections, [Wetmore et al.'s \(2015\)](#) ROC analysis led them to conclude that underlying discriminability was better for the lineup than for the showup. And this misleading conclusion is extremely consequential for our theoretical understanding of how lineups actually work. In their article, Wetmore et al. claimed that their results supported a recently-proposed “diagnostic-feature detection” theory that “presenting faces simultaneously enhances one's ability to tell the difference between innocent and guilty suspects compared to when they are presented in isolation” ([Wixted & Mickes, 2014, p. 269](#)). Specifically, Wetmore et al. reasoned that “better discriminability occurs in lineups because multiple lineup members can be compared. This allows the diagnostic features to receive more attention. A showup does not allow this comparison, and consequently, diagnostic features never become apparent” (p.13). But the appropriate  $3 \times 2$  representation of the lineup data from Wetmore et al. provides no support for this diagnostic-feature detection notion. Witnesses made far more false-positive errors in the culprit-absent lineup (64.6%) than in the culprit-absent showup (42.0%) and there were no significant differences in accurate identifications between the culprit-present lineup and showup.

### 2. The action is in filler siphoning, not improved discriminability

If underlying discriminability in lineups is not better than in showups, why has the eyewitness literature long concluded that lineups are better than showups? The answer has been known for quite some time. Lineups are better than showups because the task structure of lineups permits false-positive errors to be distributed to fillers whereas all false-positive errors in a showup load up on the innocent suspect ([Wells, 2001](#)). This phenomenon, which we call “filler siphoning,” is closely related to what has been called “filler shift” ([Wells, 1993](#)).

The data in [Table 1](#) clearly show that filler-siphoning, rather than improved discriminability, accounts for the entire drop in mistaken identifications of the innocent suspect when using lineups rather than showups in the [Wetmore et al. \(2015\)](#) data. Notice that innocent suspect identifications are 42% for the showup (lower left cell in Panel C) and only 10.2% for the lineup (lower left cell in Panel A). But, where did the innocent suspect identifications go? Panel A shows that the innocent suspect identifications from the showup did not shift to correct rejections when fillers were added to make a lineup; in fact, correct rejections were actually reduced by adding fillers. Instead, innocent suspect identifications in the showup simply shifted to false positives on the fillers in the lineup. That is not improved discriminability; it is simply shifting from one false positive to a different false positive. But ROC analysis cannot see this pattern when it treats false positives on fillers as if they were rejections. This approach has led ROC proponents to think that the lineup/showup difference represents improved underlying discriminability and erroneously profess support for a theory that lineups are better than showups because lineups permits witnesses to compare lineup members so as to determine which facial features are diagnostic, which facilitates better discrimination between the guilty and the innocent.

The fact that fillers are known by the legal system to be innocent (and, hence, are not prosecuted) has nothing to do with underlying discriminability. A false-positive identification of a filler in a culprit-absent lineup indicates poor discriminability in exactly the same

**Table 1**  
Lineup and showup data from Wetmore et al. (2015) (fair/immediate conditions).

Panel A. The actual 3 × 2 lineup data from Wetmore et al.			
	ID suspect	ID filler	Rejections
Culprit present	68.3% (true positives; accurate identifications of culprit)	10.0% (false positives; mistaken IDs of innocent fillers)	21.7% (false rejections)
Culprit absent	10.2% (false positives; mistaken IDs of the innocent suspect)	54.2% (false positives; mistaken IDs of innocent fillers)	35.6% (correct rejections)
Panel B. Lineup data from Wetmore et al. <b>after</b> being forced into a 2 × 2 for purposes of ROC analysis			
	ID suspect	Rejection	
Culprit present	68.3%	31.7% (actually, this 31.7% includes both false rejections and false positive IDs of fillers; see Panel A)	
Culprit absent	10.2%	89.8% (actually, this 89.8% includes both correct rejections and false positive IDs of fillers; see Panel A)	
Panel C. Showup data from Wetmore et al.			
	ID suspect	Rejection	
Culprit present	62.1% (true positives; accurate identifications of culprit)	37.9% (false rejections)	
Culprit absent	42.0% (false positives; mistaken IDs of the innocent suspect)	58.0% (correct rejections)	

way and to the same extent as does a false-positive identification of the innocent suspect; both are false-positive responses to noise alone.

As long as ROC analysis of lineups treats false-positive identifications of fillers as if they were rejections, ROC analyses can never claim to measure discriminability and will instead confuse filler siphoning with discriminability. To further illustrate our point, we show in Table 2 two different distributions that would yield identical ROC curves. Clearly, the data in Panel A reflect relatively high discriminability. But the data in Panel B indicate poor discriminability because in the culprit-absent condition 99% of decisions are false-positive identification errors. Nevertheless, ROC analysis would have us believe that discriminability is the same in Panels A and B of Table 2.

### 3. Filler siphoning depends on relative similarity

Of course, filler siphoning occurs only to the extent that the fillers are competitive with the suspect in terms of similarity to the eyewitness's memory of the culprit. This explains why fair lineups are more protective of innocent suspects than are biased lineups and also why filler siphoning is strongly mitigated in culprit-present lineups.

In a fair culprit-absent lineup (i.e., the innocent suspect is no more similar to the culprit than are the fillers), good fillers will siphon choices away from the innocent suspect because the fillers effectively compete with the innocent suspect for similarity to the culprit. The Wetmore et al. (2015) data show that their fair lineup resulted in the innocent suspect receiving 10.2% mistaken identifications and each of the five fillers averaged 10.8% mistaken identifications. Hence, the average filler was competitive with the

innocent suspect in terms of similarity to the culprit. If a lineup is biased, however, filler siphoning is less effective in drawing away these mistaken identifications. Wetmore et al. had a condition in which the lineup was biased and it showed approximately the same overall rate of choosing in the culprit-absent lineup as did the fair lineup, but the innocent suspect received 28.1% mistaken identifications and the average filler received only 6.3% mistaken identifications. Clearly, the fillers did not resemble the culprit as much as did the innocent suspect—the defining feature of a biased lineup. Note that the primary process that made the fair lineup more protective of the innocent suspect (10.2% mistaken identifications of the innocent suspect) than the biased lineup (28.1% mistaken identifications of the innocent suspect) was the filler-siphoning effect, not an increase in correct rejections (in fact, correct rejections were slightly but non-significantly higher for the biased lineup). The active mechanism here is filler siphoning, not underlying discriminability, but that mechanism is completely obscured by ROC analysis.

Relative similarity also explains why good fillers tend not to siphon identifications away from the culprit in the culprit-present lineup as much as they siphon identifications away from the innocent suspect in a culprit-absent lineup. Specifically, good fillers in a culprit-present lineup do not resemble the culprit as much as the culprit resembles himself and, therefore, adding good fillers to create a fair culprit-present lineup has much less effect on identifications of the culprit than adding good fillers to create a fair culprit-absent lineup has on identifications of an innocent suspect. In other words, relative similarity explains why filler siphoning is strongly muted in the culprit-present condition relative to the culprit-absent condition.

A critic might argue that filler siphoning did not occur in the culprit-present lineup in the Wetmore et al. (2015) data because

**Table 2**  
Two different lineup outcomes that would produce identical ROC Curves and lead one to conclude that discriminability is identical.

Panel A			
	ID suspect	ID filler	Rejection
Culprit present	68% (true positives)	1% (false positives on innocent fillers)	31% (false rejections)
Culprit absent	10% (false positives on innocent suspect)	1% (false positives on innocent fillers)	89% (correct rejections)
Panel B			
	ID suspect	ID filler	Rejection
Culprit present	68% (true positives)	31% (false positives on innocent fillers)	1% (false rejections)
Culprit absent	10% (false positives on innocent suspect)	89% (false positives on innocent fillers)	1% (correct rejections)

the accurate identification rate was higher in the lineup than in the showup (albeit only by a small and a non-significant 6%). But that ignores the fact that the choosing rate was 16.2% higher for the culprit-present lineup (78.3% choosing) than for the culprit-present showup (62.1% choosing), and the vast majority of this increase in choosing (10% of the 16.2%) went to fillers. Hence, even though the fillers in the culprit-present lineup did not siphon identifications away from the culprit, the fillers still distributed the majority of additional choices from the lineup to the fillers instead of to the culprit. Because ROC analysis collapses filler identifications with rejections, ROC is blind to the fact that choosing rates are higher with lineups than showups and to the regularities associated with filler siphoning.

#### 4. Computational modeling of the problem

Although we believe our explanation of why ROC analysis does not measure underlying discriminability is compelling, recent computational modeling provides additional support for our argument. Lampinen (2015) used computational modeling to generate 40,000 data points per experiment comparing ROC curves on showups to ROC curves on lineups. In each experiment Lampinen set discriminability and decision criteria to be identical for lineups and showups. Recall that showups are  $2 \times 2$  task structures that perfectly fit the Signal Detection model on which ROC is based. Hence, if ROC analyses of lineups measure discriminability then the ROC curves for lineups should overlap the ROC curves for showups. In fact, however, ROC curves for lineups diverged substantially and consistently from the ROC curves for showups even though underlying discriminability and decision criterion were the same. The divergence of ROC curves held across a variety of levels of discriminability, decision strategies (best match versus a mixed best-match/relative match), equal variance, and unequal variance models. Lampinen's computational modeling is strong additional evidence that ROC analyses on lineups are not measures of discriminability.

#### 5. The fallback position for ROC proponents and the Bayesian approach

The “luster” of the ROC approach to lineups derives from the assumption that the area under the ROC curve measures underlying discriminability. We have shown here how and why that assumption is false. So, if ROC analysis on lineups is not a measure of discriminability, what is it? A fallback claim by ROC proponents might be that ROC curves on lineups are measures of *applied* utility (rather than discriminability). Consider again the two panels of Table 2. Although discriminability should not be construed as the same in the two panels, we could conclude that the outcomes for the guilty and innocent suspect are identical in both. In other words, the guilty suspect has a 68% chance of being accurately identified and the innocent suspect has a 10% risk of being mistakenly identified in both panels.

If ROC proponents fall back on this applied utility argument (abandoning the claim of discriminability), however, then the ROC approach has some stiff competition, especially from the Bayesian approach to lineups, which was launched 35 years ago (Wells & Lindsay, 1980). Space does not permit a full treatment of the Bayesian approach (see Wells et al., for a full treatment), but the Bayesian approach has several advantages over the ROC approach. First, a Bayesian analysis of lineup data does not require that filler identifications be treated as though they were rejections. Instead, any  $3 \times 2$  lineup data array (like that in Table 1, Panel A) generates three Bayesian curves, one for suspect identifications, one for filler identifications, and one for rejections. Hence, with the Bayesian

approach there is no need to treat a witness behavior (such as a positive identification of a filler) as though it were something other than what it actually is. Second, the Bayesian approach necessarily and efficiently takes into consideration the base rate for the culprit-present versus culprit-absent variable, which has a huge impact on applied outcomes. Whereas an ROC analysis assumes a 50/50 base rate, a Bayesian analysis generates curves that examine posterior probabilities that the suspect is guilty across the entire range of possible base rates. Third, different types of Bayesian curves have been developed that can answer different questions about eyewitness identification evidence. For example, *prior-by-posterior* curves reveal how much one can trust an identification of the suspect as evidence of guilt or how much one can trust a rejection as evidence of innocence. *Information-gain* curves tell us how much we should revise our belief about guilt of the suspect (upward or downward) as a function of whether the witness identified the suspect, identified a filler, or rejected the lineup.

The Bayesian approach has been a strong discovery tool for lineups (Wells & Lindsay, 1980; Wells & Turtle, 1986; Wells & Olson, 2002; Wells et al., 2015a,b). For example, Bayesian analyses show that one lineup procedure might be more diagnostic for identifications of the suspect (more indicative of guilt) whereas another procedure might be more diagnostic for rejections (more indicative of innocence). In other words, the Bayesian procedure (but not the ROC approach) can show one lineup procedure to be better at incriminating the guilty (when the witness makes an identification of the suspect) whereas another procedure is better at exculpating the innocent (when the witness makes a rejection decision; see Wells et al., 2015a,b). The ROC approach, in contrast, is focused exclusively on the incriminating value of a particular lineup procedure (focusing on only two of the six cells in lineups). This is an important shortcoming because the applied utility of a lineup procedure is not restricted to its ability to incriminate; lineups also have potential exculpatory value. Good lineup measures should be able to quantify not only the incriminating value of a lineup procedure but also its exculpatory value, which the Bayesian approach does but the ROC approach does not.

#### 6. Misplaced criticisms of the Bayesian approach

ROC proponents have argued that measures of “probative value” should never be used because such measures do not control for decision criterion (e.g., see Wixted & Mickes, 2012). But proponents of the probative value approach never argued that they were controlling for decision criterion. Probative value is a longstanding legal concept of relevance to courts that refers to the extent to which a piece of evidence changes the likelihood of some proposition. Pryke, Lindsay, Dysart, and Dupuis (2004) defined probative value in terms of the probability that a suspect is guilty given an identification of the suspect (what proportion of suspect identifications are of the guilty suspect?) whereas Clark and Godfrey (2009) defined probative value in terms of the proportion of suspect identifications that are of the innocent suspect (i.e., 1-proportion guilty, or innocence risk). Both of these have mathematical equivalence to the likelihood ratio, which is simply the ratio of identifications of the guilty suspect to identifications of the innocent suspect (Wells & Lindsay, 1980). It is important to note that these indexes of probative value were not pulled out of thin air. In fact, each of the probative value measures are simple derivations from Bayes Theorem.

The probative-value approach takes the data for what they are rather than what they might be if we controlled for decision criterion. Nevertheless, overlooked by ROC proponents is the fact that the Bayesian approach reveals shifts in decision criterion as long as researchers look at the probative value of all three possible witness responses (suspect identifications, filler identifications

and rejections) as recently explained by Wells et al. (2015b). Specifically if probative value for an identification of the suspect for Procedure A is higher than for Procedure B due solely to a conservative criterion shift, then there will be a cost in probative value for rejections (i.e., rejections will be more probative for procedure B).

As we argued earlier, we do not believe that ROC analyses on lineups are measures of discriminability. For similar reasons, we are also not convinced that ROC analyses on lineups accurately portray differences in decision criterion. Consider again the hypothetical data in Panels A and B of Table 2. In Panel A, responding is quite conservative as witnesses are explicitly rejecting the lineup 60% of the time (average over culprit present and culprit absent) whereas in Panel B witnesses are responding liberally by rejecting the lineup only 1% of the time. But ROC analysis would lead us to believe that decision criteria are equivalent in Panels A and B in Table 2, because the suspect is chosen 39% of the time (averaging over culprit present and culprit absent) in both procedures. Hence, changes in decision criterion are not accurately tracked over the ROC space.

ROC proponents have also argued that there is value in the fact that ROC curves display data at each level of witness confidence (cumulatively from higher levels of confidence to lower levels of confidence). We agree that there is value in examining data at each level of confidence. But this idea is hardly new or unique to ROC analysis. Examining eyewitness-identification accuracy at each level of confidence made its appearance in the eyewitness literature nearly 30 years ago (Lindsay, 1986) and is a central feature of calibration analyses of lineups, which date back nearly 20 years (Juslin, Olson, & Winman, 1996). More recent work shows how the Bayesian approach can examine identification performance at each level of confidence and, unlike the ROC approach, can do this separately for identifications of the suspect, filler identifications, and rejections (Wells et al., 2015b).

## 7. Concluding remarks

ROC analysis is a useful tool for the  $2 \times 2$  problem structure for which it was devised. But lineups are a natural  $3 \times 2$  problem structure. Forcing the  $3 \times 2$  lineup structure into a  $2 \times 2$  structure by treating filler identifications as if they were rejections, rather than the false-positive errors that they are, violates a fundamental assumption of Signal Detection Theory that false-positive errors are false-positive errors and not correct rejections. As a result, ROC curves on lineups are not measures of discriminability. Our logical arguments for this are further supported by Lampinen's (2015) computational modeling in which he showed that a  $2 \times 2$  showup task produced a different ROC curve than a  $3 \times 2$  lineup task even though discriminability and decision criterion were held constant. Moreover, collapsing the natural  $3 \times 2$  lineup structure into a  $2 \times 2$  structure obscures what is actually happening in the data with regard to filler siphoning effects. This, in turn, leads to theoretical confusion regarding how lineups work.

We do not believe that any single method is the *only* way to evaluate lineups. But whatever method is used must treat the lineup as the  $3 \times 2$  problem that it is and not treat fillers as though they were rejections.

## Conflict of interest statement

The authors declare that they have no conflict of interest.

## Acknowledgements

Aspects of this work were supported by grant SES0850401 from the National Science Foundation to the lead author.

## References

- Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review*, 16, 22–42. <http://dx.doi.org/10.3758/PBR.16.1.22>
- Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using ROC analysis. *Current Directions in Psychological Science*, 23, 3–10. <http://dx.doi.org/10.1177/0963721413498891>
- Juslin, P., Olson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: comments on what can and cannot be inferred from a low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 5, 1304–1316. <http://dx.doi.org/10.1037/0278-7393.22.5.1304>
- Lampinen, J. M. (2015). *ROC analyses in eyewitness identification research*. (in press).
- Lindsay, R. C. L. (1986). Confidence and accuracy of eyewitness identification from lineups. *Law & Human Behavior*, 10, 229–239. <http://dx.doi.org/10.1007/BF01046212>
- 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. <http://dx.doi.org/10.1016/j.jarmac.2014.04.007>
- 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. <http://dx.doi.org/10.1037/a0030609>
- Mickes, L., & Wixted, J. T. (2015). On the applied implications of the “verbal overshadowing effect”. *Psychological Science*, 10, 400–403. <http://dx.doi.org/10.1177/1745691615576762>
- Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: a method of calibrating eyewitness identifications. *Journal of Applied Psychology*, 89, 73–84. <http://dx.doi.org/10.1037/0021-9010.89.1.73>
- Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, 48, 553–571. <http://dx.doi.org/10.1037/0003-066X.48.5.553>
- Wells, G. L. (2001). Eyewitness lineups: data, theory, and policy. *Psychology, Public Policy, and Law*, 7, 791–801. <http://dx.doi.org/10.1037/1076-8971.7.4.791>
- Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, 88, 776–784. <http://dx.doi.org/10.1037/0033-2909.88.3.776>
- Wells, G. L., & Olson, E. (2002). Eyewitness identification: information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, 8, 155–167. <http://dx.doi.org/10.1037/1076-898X.8.3.155>
- Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: an experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, 39, 1–14. <http://dx.doi.org/10.1037/lhb0000096>
- Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: the importance of lineup models. *Psychological Bulletin*, 99, 320–329. <http://dx.doi.org/10.1037/0033-2909.99.3.320>
- 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, 99–102. <http://dx.doi.org/10.1037/lhb0000125>
- Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, 4, 8–14. <http://dx.doi.org/10.1016/j.jarmac.2014.07.003>
- Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science*, 7, 275–278. <http://dx.doi.org/10.1177/1745691612442906>
- Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature model of eyewitness identification. *Psychological Review*, 121, 262–276. <http://dx.doi.org/10.1037/a0035940>