Journal of Applied Research in Memory and Cognition 4 (2015) 329–334

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Reply

ROC analysis measures objective discriminability for any eyewitness identification procedure☆

John T. Wixted[a,*], Laura Mickes[b]

[a] Department of Psychology, University of California, San Diego, United States
[b] Department of Psychology, Royal Holloway, University of London, United Kingdom

ARTICLE INFO

Article history:
Received 24 August 2015
Accepted 30 August 2015
Available online 10 November 2015

Keywords:
Eyewitness identification
ROC analysis
Discriminability
Bayesian analysis
Filler siphoning

ABSTRACT

Which eyewitness identification procedure better enables eyewitnesses to discriminate between innocent and guilty suspects? In other words, which procedure better enables eyewitnesses to sort innocent and guilty suspects into their correct categories? The answer to that objective, theory-free question is what policymakers need to know, and it is precisely the information that ROC analysis provides. Wells et al. largely ignore that question and focus instead on whether ROC analysis accurately measures underlying (theoretical) discriminability for lineups. They argue that the apparent discriminability advantage for lineups over showups is an illusion caused by "filler siphoning." Here, we demonstrate that, both objectively and theoretically, the ability of eyewitnesses to discriminate innocent from guilty suspects is higher for lineups compared to showups, just as the ROC data suggest. Intuitions notwithstanding, filler siphoning does not account for the discriminability advantage for lineups. An actual theory of discriminability is needed to explain that interesting phenomenon.

The two competing claims in this debate could not be clearer:

1. Wells, Smalarz, and Smith (2015) claim that ROC analysis does not measure discriminability when lineups are used (because lineups have fillers) and so cannot be used to evaluate the diagnostic accuracy of that eyewitness identification procedure; instead, in their view, a Bayesian analysis, based on the joint consideration of the diagnosticity ratio and base rates, offers a better way to measure the diagnostic accuracy of lineups.
2. We claim that ROC analysis *does* measure discriminability when lineups are used (despite the presence of fillers) and is the only definitive way to measure diagnostic accuracy; moreover, just as in diagnostic medicine, a Bayesian analysis has *no bearing whatsoever* on the diagnostic accuracy of a lineup procedure.

It is hard to imagine a more urgent issue for the field to resolve because only one of these arguments can be correct, yet both approaches (ROC analysis and Bayesian analysis) are being used to adjudicate important applied questions, such as whether or not simultaneous lineups are diagnostically superior to sequential lineups. A National Academy of Sciences committee on eyewitness identification recently endorsed ROC analysis over the longstanding Bayesian approach based on the diagnosticity ratio (National Research Council, 2014). We believe they made the right call.

As shown in Fig. 1 the fair lineup condition from Wetmore et al. (2015) yielded a higher ROC curve (i.e., higher discriminability) than the showup condition. Wells et al. (2015) used the overall correct and false ID rates from those two conditions to illustrate their claim that, theoretically, lineups do not yield higher discriminability than showups – contrary to what ROC analysis suggests.

We agree with Wells et al. (2015) that the Wetmore et al. (2015) data can be used to conclusively settle the debate about what ROC analysis actually measures, so we focus much of our response on those data. We first consider objective (theory-free) discriminability, which is the only concern of policymakers. We then focus on theoretical discriminability, which is of concern to theoreticians (not policymakers) yet was the main focus of the Wells et al. critique.

1. Objective discriminability

Imagine a group of 100 innocent and 100 guilty suspects. Wetmore et al. (2015) found that the overall correct ID rate for the showup procedure was .61. Thus, using a showup, 61 out of the

☆ This work was supported in part by the National Science Foundation [SES-1456571] to John T. Wixted and the Economic and Social Research Council [ES/L012642/1] to Laura Mickes and John T. Wixted. The content is solely the responsibility of the authors and does not necessarily reflect the views of the National Science Foundation or the Economic and Social Research Council.

* Corresponding author. Tel.: +1 858 534 3956.
E-mail address: jwixted@ucsd.edu (J.T. Wixted).

http://dx.doi.org/10.1016/j.jarmac.2015.08.007
2211-3681/

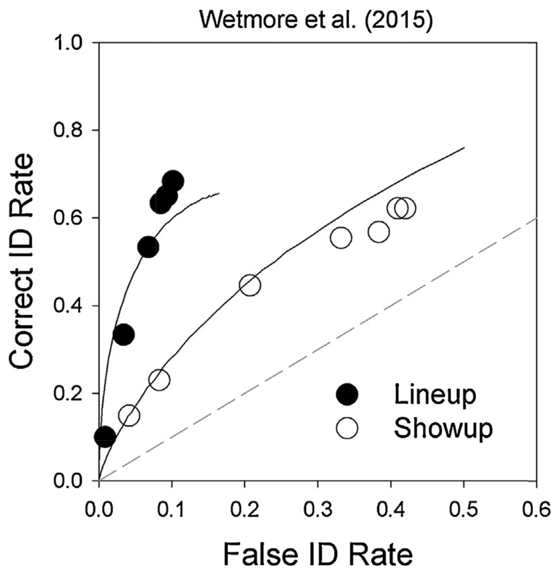

Fig. 1. Lineup and showup ROC data from Wetmore et al. (2015). The smooth curves indicate fits to a theoretical model, which is discussed in a later section entitled "Underlying (Theoretical) Discriminability." The rightmost point on each ROC represents the overall correct and false ID rates. The dashed gray line indicates chance performance.

100 guilty suspects would be correctly classified as guilty. The overall false ID rate for the showup procedure was .42. Thus, using the showup, 42 out of the 100 innocent suspects would be incorrectly classified as guilty. For the lineup, the correct and false ID rates were .67 and .10, respectively, so 67 out of the 100 guilty suspects would be correctly classified as guilty and 10 of the 100 innocent suspects would be incorrectly classified as guilty. Thus, using the lineup, more of the 100 innocent suspects *and* more of the 100 guilty suspects would be correctly classified. No theoretical model – and no consideration of filler IDs – is needed to appreciate the fact that the lineup yields higher objective discriminability in that it more accurately classifies both innocent and guilty suspects than the showup. When the question concerns which procedure more accurately discriminates innocent from guilty suspects, filler IDs are simply irrelevant.

The superiority of the lineup would not change if the base rates of innocent and guilty suspects were no longer equal. Imagine, for example, a mixture of 100 guilty suspects and 1000 innocent suspects. Using the showup, 61 of the 100 guilty suspects would be correctly classified as guilty, and 420 of the 1000 innocent suspects would be incorrectly classified as guilty. Using the lineup, 67 of the 100 guilty suspects would be correctly classified as guilty, but only 100 of the 1000 innocent suspects would be incorrectly classified as guilty. Thus, no matter what the base rates, the lineup procedure more accurately classifies both innocent and guilty suspects than the showup does. This example illustrates the fact that Bayesian considerations play *no role whatsoever* in determining which eyewitness identification procedure better classifies innocent and guilty suspects into their proper categories.

ROC analysis measures objective discriminability. What does all of this have to do with ROC analysis? Wells et al. (2015) focused on the overall correct and false ID rates from each procedure used by Wetmore et al. (2015); namely, the rightmost ROC point for each procedure, but the same logic applies to all of the correct and false ID rates that can be achieved by either eyewitness identification procedure. A 6-person fair lineup can achieve false ID rates in the range of 0–.167 (because always choosing from a fair target-absent lineup would result in the innocent suspect being identified 1/6 = .167 of the time). In that range, consider the achievable correct ID rates for the showup in Fig. 1 (indicated by the smooth curve) and choose the point that, according to your subjective values, most appropriately balances the costs of a false ID and the benefits of a correct ID. No matter which showup point you pick, now consider the fact that the lineup – because it yields a higher ROC – can achieve a higher correct ID rate and, at the same time, a lower false ID rate than your preferred showup point. The fact that the lineup can achieve a superior outcome remains true no matter what the base rates of target-present and target-absent lineups might be and no matter what the filler ID rate might be for the lineup. As a general rule (not just for the Wetmore et al. data), the procedure that yields the higher ROC can simultaneously achieve a higher correct ID rate *and* lower false ID rate than the procedure that yields the lower ROC. That is precisely why the procedure that yields a higher ROC is, objectively (and, we would add, unarguably), the diagnostically superior procedure.

ROC analysis vs. Bayesian analysis. Wells et al. (2015) mistakenly assert that ". . . ROC analysis assumes a 50/50 base rate. . .", but each ROC point is independent of the base rate, just as the overall correct and false ID rates are. They also say that ". . .a Bayesian analysis generates curves that examine posterior probabilities that the suspect is guilty across the entire range of possible base rates." However, a Bayesian analysis merely quantifies the posterior odds of guilt for a particular suspect who has been identified by an eyewitness: posterior odds = prior odds times the diagnosticity ratio (cf. Zweig & Campbell, 1993). The posterior odds of guilt can be high or low even when using an inferior diagnostic procedure, depending on how conservative or liberal the decision criterion is. ROC analysis, by contrast, tells you which diagnostic procedure does a better job of sorting innocent and guilty suspects into their correct categories. Thus, ROC analysis and Bayesian analysis *address different questions*; they are in no way competing methods for identifying the diagnostically more accurate eyewitness identification procedure. Only ROC analysis can do that.

2. Underlying (theoretical) discriminability

Wells et al. (2015) focus mainly on theoretical discriminability even though it is not relevant to the debate over the applied utility of ROC analysis. For example, they make the following claim: "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*" (p. 7, emphasis added). Underlying discriminability is theoretical discriminability, which is of interest to theoreticians (e.g., Mickes, Wixted, & Wais, 2007; Wixted & Mickes, 2010, 2014), not to policymakers.

Wells et al. (2015) argue that theoretical discriminability is not higher for lineups than showups. In their view, the apparently higher discriminability for lineups in Fig. 1 is an illusion caused by the many filler IDs that lineups occasion. However, Wells et al. relied on intuition alone to analyze this theoretical issue. We now analyze the same data using signal-detection theory – the standard theory of recognition memory for more than half a century (Egan, 1958).

A simple signal-detection model for lineups. According to the simplest signal-detection model (Fig. 2) memory strength values for fillers, innocent suspects and guilty suspects are distributed according to Gaussian distributions with means of μ_{Filler}, $\mu_{Innocent}$, and μ_{Guilty}, respectively. A 6-member target-present lineup is conceptualized as 5 random draws from the Filler distribution and 1 random draw from the Guilty distribution; a 6-member target-absent lineup is conceptualized as 5 random draws from the Filler distribution and 1 random draw from the Innocent distribution. If a fair target-absent lineup is used, then $\mu_{Filler} = \mu_{Innocent}$, in which case the model reduces to a 2-distribution model. Of primary interest