No possibility of a Selection Bias but Direct Evidence of a Simultaneous Superiority Effect: A Reply to Wells et al.

Karen L. Amendola[1] & John T. Wixted[2]

[1]Police Foundation

[2]University of California, San Diego

RUNNING HEAD: SIMULTANEOUS SUPERIORITY, NOT A SELECTION BIAS

*Correspondence concerning this article should be addressed to Karen L. Amendola, Police Foundation, 1201 Connecticut Avenue, NW, Suite 200, Washington, DC 20036-2636. E-mail: kamendola@policefoundation.org or John Wixted, Department of Psychology, 0109, University of California, San Diego, La Jolla, CA 92093-0109. E-mail: jwixted@ucsd.edu.

Keywords: Eyewitness Identification, ROC Analysis, Sequential Lineups, Simultaneous Lineups

Comparing the Diagnostic Accuracy of Suspect Identifications made by Actual Eyewitnesses

No possibility of a Selection Bias but Direct Evidence of a Simultaneous Superiority Effect: A Reply to Wells et al.

Wells et al. (this issue) argue that the suspects from the AJS field study whose case files were rated for evidence of guilt in our Study B were strikingly imbalanced across lineup type (simultaneous vs. sequential) and prior case outcome ("adjudicated guilty" vs. "not prosecuted"). According to them, this unexplained disparity means that our study was heavily biased against the sequential procedure, but it actually means something else.

Our Study B began with 340 protocol-consistent lineups[1] administered in Austin, Texas (one of four AJS study sites). These lineups were selected because they were classified as "pristine" by Wells et al. (2011), and they are the same lineups that were analyzed by them. Based on eligibility criteria (including State Law)[2], 76 lineups had to be eliminated from consideration, all prior to knowing the case outcomes. From the remaining eligible cases, *all* of the identified suspects were included in our Study B[3]. Thus, the insinuation of biased selection – specifically, the idea that simultaneous suspects known to have been adjudicated guilty were preferentially selected for inclusion in our ratings study – can be summarily dismissed. There simply was no mechanism for case selection to be biased in favor of either lineup format, which means that some other explanation for the disparity in case outcomes must apply. That explanation turns out to be a simple and obvious one.

---

[1] Of the 615 lineups, 275 did not meet the Wells et al. criteria of "protocol consistent," resulting in 340 lineups

[2] For example, Texas state law required exclusion of cases involving juveniles, the DA's office directed us to exclude cases involving sexual assault, some cases were found to have irreconcilable inconsistencies in case details, some made no mention of the suspect, etc. (see Amendola et al., 2014).

[3] The stratified random sampling process that Wells et al. mention was applied to the selection of no pick cases, not to the selection of suspect cases.

Amendola and Wixted (2015) used two "ground truth" measures to determine the likelihood of guilt associated with suspects identified from simultaneous and sequential lineups. Given the approximately equal suspect choosing rates for the two lineup formats used in the AJS field study, the superior lineup procedure is the one that yields a higher percentage of identified suspects who are guilty and (therefore) a lower percentage who are innocent. Study A examined case outcomes of suspects identified at all 4 of the AJS study sites, whereas Study B examined expert ratings of guilt associated with suspects identified at the Austin site only. Limiting our Study B to the Austin site had the advantage of retaining the large majority of the AJS study lineups (nearly 70% of the lineups were administered at that one site) while eliminating error variance due to site differences. For the same reason (i.e., to reduce error variance, thereby increasing the odds of detecting a real effect), it would not be unreasonable to evaluate the case-outcome data limited to the Austin site. Although we did not do so in our original report, Wells et al. did exactly that in their comment. For the suspects identified in Austin, the case outcome data (Table 1) favored simultaneous lineups: 77.3% of simultaneous suspects were adjudicated guilty, but only 53.3% of sequential suspects were adjudicated guilty. This difference is not quite significant (p = .077), but Wells et al. regard it as a real effect that is in need of explanation.

What accounts for this effect? Because biased selection is not a relevant consideration, the obvious explanation for why a higher percentage of simultaneous suspects were adjudicated guilty is that simultaneous lineups support the identification of more guilty suspects – and fewer innocent suspects – than sequential lineups. Our expert ratings data (Study B) confirmed that suspects identified from simultaneous lineups had significantly more objective evidence of guilt against them compared to suspects identified from sequential lineups. Thus, according to

multiple measures, the Austin-based AJS field study data indicate that simultaneous lineups are diagnostically superior to sequential lineups.

Wells et al. did not identify a flaw in our study (what looked like an inexplicable bias to them is direct evidence of a simultaneous superiority effect), but their comment does underscore how much the terms of the debate have changed in just the last few years. Not long ago, the field debated whether or not the *sequential* superiority effect was real; now, the only question is whether or not the *simultaneous* superiority effect is real. The idea that the sequential procedure is superior to the simultaneous procedure originated from laboratory studies conducted between 1985 and 2011, which often found that the diagnosticity ratio was higher when the sequential procedure was used (Steblay, Dysart & Wells, 2011). Recently, Wixted and Mickes (2012) and Gronlund, Wixted and Mickes (2014) argued that the intuitively appealing diagnosticity ratio can be misleading and that ROC analysis is needed to determine which lineup procedure is diagnostically superior. Wells (2014) disagreed – reiterating his longstanding commitment to the diagnosticity ratio – but a new report from the National Academy of Sciences (NAS) on eyewitness identification declared that "ROC analysis represents an improvement over a single diagnosticity ratio" (National Research Council, 2014, p. 80). This is an important conclusion by the prestigious NAS committee because, without exception, laboratory-based studies that have been interpreted as favoring the sequential procedure relied on single diagnosticity ratios to assess lineup performance (Steblay et al., 2011). If an individual diagnosticity ratio cannot identify the superior lineup procedure, it follows that the entire body of research claiming a sequential superiority effect is now in question. Indeed, because they were all based on single diagnosticity ratios, there is not a single laboratory-based study that has ever reported convincing evidence of a sequential superiority effect based on what we now know. By contrast, recent

laboratory studies using ROC analysis have consistently found evidence in favor of the simultaneous procedure (e.g., Mickes, Flowe & Wixted, 2012). Our Phase 2 analysis of the AJS field study adds real-world validity to this increasing body of laboratory-based evidence pointing to a simultaneous superiority effect, although we agree with the NAS committee that more research is needed.

Findings from another field study comparing simultaneous and sequential lineups administered by the Houston Police Department were just released (W. Wells, 2014). For comparative purposes, Table 2A presents final eyewitness decisions from the Phase 1 analysis of the AJS field study (Wells, Steblay & Dysart, 2014), and Table 2B reports the corresponding results from the comparable (blind) conditions of the Houston field study. Neither study found a difference in suspect choosing rates associated with the two lineup procedures, but both found non-significant trends with regard to filler IDs. G. Wells et al. (2014) interpreted the non-significant trend towards higher filler IDs for simultaneous lineups (Table 2A) study to favor sequential lineups, but the new Houston field study exhibits a similar non-significant trend in the opposite direction (Table 2B). Even if one assumes that filler IDs can be used to identify the better procedure (which does not seem to be the case at all), the combined evidence from these two randomized field trials provides no evidence that sequential lineups are superior. Thus, after 30 years of research, convincing evidence for a sequential superiority effect is nowhere to be found, either in studies conducted in the lab (which relied on the diagnosticity ratio) or in studies conducted in the field (based on filler IDs). By contrast, a growing body of recent evidence, both in the lab (using ROC analysis) and in the field (based on evidentiary strength ratings in our Study B), points to a simultaneous superiority effect.

References

Amendola, K.L., Valdovinos, M.D., Hamilton, E.E., Slipka, M.G., Sigler, M., and Kaufman, A. (2014). Photo arrays in eyewitness identification procedures: Presentation methods, influence of ID decisions on experts' evaluations of evidentiary strength, and follow-up on the AJS Eyewitness ID Field Study. Washington, DC: Police Foundation http://www.policefoundation.org/sites/glfiles/g798246/f/201403/FINAL%20EWID%20RBPORT-Police%20Foundation%281%29-1_O.pdf

Gronlund, S. D., Wixted, J. T. & Mickes, L. (2014). Evaluating eyewitness identification procedures using ROC analysis. *Current Directions in Psychological Science, 23*, 3-10.

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.

National Research Council (2014). *Identifying the Culprit: Assessing Eyewitness Identification*. Washington, DC: The National Academies Press.

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17,* 99-139.

Wells, G. L. (2014). Eyewitness Identification: Probative Value, Criterion Shifts, and Policy Regarding the Sequential Lineup. *Current Directions In Psychological Science, 23*, 11-16.

Wells, G.L., Steblay, N.K., & Dysart, J.E. (2014). Double-Blind Photo-Lineups Using Actual Eyewitnesses: An Experimental Test of a Sequential versus Simultaneous Lineup Procedure. *Law and Human Behavior*.

Wells, W. (2014). The Houston Police Department Eyewitness Identification Experiment: Analysis and Results. Retrieved from: http://www.lemitonline.org/research/projects.html

Table 1

Observed distribution of suspects across lineup type, simultaneous (SIM) vs. sequential (SEQ), and case outcome, adjudicated guilty vs. not prosecuted.

Case Outcome	SIM	SEQ
Adjudicated Guilty	17	16
Not Prosecuted	5	14
% Adjudicated Guilty	77.3%	53.3%

Note: Case outcome information was unavailable for 5 of the 30 sequential suspects, but Wells et al. placed all 5 into the "not prosecuted" category. It is not unreasonable to assume that these 5 suspects were not prosecuted (e.g., perhaps the DA's office and Police Department did not retain information about those cases for that reason), so we accept that assumption.

Table 2

Percentage of Participants who Picked a Suspect, Picked a Filler, or Rejected the Lineup when Simultaneous (SIM) or Sequential (SEQ) Lineups were used in the AJS Field Study (A) and in the comparable ("blind") conditions of the Houston Field Study (B).

A

Category	SIM	SEQ	*p*
No ID	0.57	0.61	0.39
Suspect ID	0.25	0.27	0.67
Filler ID	0.18	0.12	0.09

B

Category	SIM	SEQ	*p*
No ID	0.38	0.36	0.77
Suspect ID	0.36	0.29	0.15
Filler ID	0.26	0.35	0.07

Note: The *p*-value of the statistical test comparing the difference in the proportion of IDs for SIM vs. SEQ (*p*) has not been adjusted for multiple comparisons.