Article in press: Psychology, Public Policy, and Law

Seventy-Two Tests of the Sequential Lineup Superiority Effect: A Meta-Analysis and Policy Discussion

Nancy K. Steblay

Augsburg College

Jennifer E. Dysart

John Jay College of Criminal Justice

Gary L. Wells

Iowa State University

Author Note

Nancy K. Steblay, Department of Psychology, Augsburg College; Jennifer E. Dysart, Department of Psychology, John Jay College of Criminal Justice; Gary L. Wells, Department of Psychology, Iowa State University.

Data collection and analysis was supported by Award No. 2007-IJ-CX-0046 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

Correspondence concerning this article should be addressed to Nancy K. Steblay, Department of Psychology, Augsburg College, 2211 Riverside Avenue, Minneapolis, MN 55454. E-mail: steblay@augsburg.edu

Abstract

Nearly a decade ago, a meta-analysis showed that identification of a suspect from a sequential versus a simultaneous lineup was more diagnostic of guilt (Steblay, Dysart, Fulero, & Lindsay, 2001). Since then, controversy and debate regarding sequential-superiority has emerged. We report the results of a new meta-analysis involving 72 tests of simultaneous and sequential lineups from 23 different labs involving 13,143 participant-witnesses. The results are very similar to the 2001 results in showing that the sequential lineup is less likely to result in an identification of the suspect, but also more diagnostic of guilt than the simultaneous lineup. An examination of the *full diagnostic design dataset* (27 tests that used the full simultaneous/ sequential x culprit-present/culprit-absent design) showed that the average gap in correct identifications favoring the simultaneous lineup over the sequential lineup—8%— is smaller than the 15% figure obtained from the 2001 meta-analysis (and from the current full 72-test dataset). The lower error rate incurred for culprit-absent lineups with use of a sequential format remains consistent across the years, with 22% fewer errors than simultaneous lineups. A Bayesian analysis shows that the posterior probability of guilt following an identification of the suspect is higher for the sequential across the entire base rate for culprit presence/absence. New ways to think about policy issues are discussed.

Seventy-Two Tests of the Sequential Lineup Superiority Effect: A Meta-Analysis and Policy Discussion

The problem of wrongful conviction has drawn attention to a decades-long research effort by eyewitness scientists concerning the conditions under which eyewitness memory is less or more reliable. Experimental findings have led to specific recommendations for improvements in eyewitness evidence collection procedures, particularly for police lineups (e.g., Technical Working Group for Eyewitness Evidence, 1999; Wells, Malpass, Lindsay, Turtle, & Fulero, 2000). One recommendation is that all lineups, photographic or live, be presented to the witness one member at a time (sequentially) rather than in the traditional all-at-once (simultaneous) format (Wells, 2006).

A comparison of sequential and simultaneous lineups was published nine years ago, in a meta-analytic review of 23 experimental reports available at that time (Steblay, Dysart, Fulero, & Lindsay, 2001). The research question was whether eyewitness accuracy is affected by the display format of the lineup, and the meta-analysis established that adult witnesses who view a lineup of sequential format are significantly less likely to make a pick from the lineup compared to witnesses who view a simultaneous lineup. This difference in choosing rate translated into two key outcomes for identification accuracy. First, when the culprit was present in the lineup, the simultaneous lineup produced significantly more correct identifications of the offender (r = .14). Second, when the culprit was not in the array and thus any pick from the lineup a mistake, the sequential lineup produced significantly fewer mistaken identifications (r = .24). Likewise, a subset of nine tests in which the culprit was replaced with a similar-appearing person yielded significantly fewer false identifications of this designated innocent suspect from the sequential compared to the simultaneous lineup (*r* = .23). The results prompted the authors to claim a *sequential-superiority effect*.

In the following years further testing of the sequential lineup has occurred, and the number of sequential-simultaneous comparisons has more than doubled to 72 experimental tests. The past decade also has seen the movement of double-blind sequential lineups into police practice and law enforcement policy (e.g., Gaertner & Harrington 2009; Klobuchar, Steblay, & Caligiuri, 2006). The sequential lineup continues to draw interest, approval, and, sometimes fire, therefore examination of the extant data is a timely endeavor. Beginning with a refined definition of lineup "superiority", the current analyses focus on four issues relevant to the eyewitness research community and to policy-makers: a reassessment of the *sequential-superiority effect* in light of new data; a closer examination of the operational specifics of the sequential procedure; the comparative diagnosticity (cost-benefit ratio) of sequential and simultaneous lineups; and policy implications of the findings.

Defining Superiority

Lineup superiority is defined in this meta-analysis as a higher diagnosticity ratio (more simply called *diagnosticity*; Wells & Lindsay, 1980; Wells & Turtle, 1986). Diagnosticity indicates how much more likely one event is relative to another; in the case of eyewitness identification, this ratio reflects identifications of the culprit to identifications of an innocent suspect. Lineup performance can be evaluated by computing diagnosticity for each of the two lineup formats, simultaneous and sequential. Then, given any two diagnosticity ratios, the higher of the two is stronger evidence for the proposition that the suspect is the culprit. In the legal system, diagnosticity is known as the index of probative value, the tendency to prove or disprove the truth of an allegation. Therefore, greater lineup diagnosticity is a particularly useful index of superiority as it indicates that a witness's decision stemming from a specific lineup format is more probative of guilt.

Calculation of diagnosticity. A properly-constructed lineup includes only one suspect. If the suspect is in fact the culprit (culprit-present lineup), then an innocent suspect cannot be identified from that lineup (the other lineup members are known-innocent fillers). Likewise, if the suspect is in fact innocent (culprit-absent lineup), then an accurate identification of the culprit is not possible. Accordingly, diagnosticity is calculated as the ratio of identifications of the criminal from culprit-present lineups to identifications of the innocent suspect from culprit-absent lineups. For eyewitness identification studies in which there is no a priori designated innocent suspect in the culprit-absent lineup, the rate of mistaken identification of an innocent suspect is estimated from the choosing rate (choosing rate divided by the number of lineup members). For example, if 50% of witnesses pick the culprit from a culprit-present lineup and the choosing rate in the (6-person) culprit-absent lineup is 30%, then the diagnosticity ratio would be 50% ÷ 5% = 10.0.

Diagnosticity in the meta-analytic dataset. The calculation of diagnosticity for a comparison of sequential and simultaneous lineups relies on the assumption that conditions for testing culprit-present lineups and culprit-absent lineups are matched within each study. In other words, features of the crime event, lineup stimuli, participant-witness populations, and all other factors must be held constant between culprit-present and culprit-absent lineups and between simultaneous and sequential lineups. Only studies that included the full 2 (culprit present/absent) X 2 (simultaneous/ sequential) design using a fully randomized procedure can meet this important criterion. If studies that did not use the fully randomized 2 X 2 design were included, then the diagnosticity ratio would be untrustworthy because it would involve comparisons of simultaneous lineups from one study against sequential lineups from another study or comparisons of the culprit-present conditions of one study to the culprit-absent conditions of another study. These studies would almost always differ in other ways, thereby not permitting a full-design estimate of the diagnosticity ratio.

Posterior (after identification) probabilities of guilt. Diagnosticity is important for another reason. The diagnosticity ratio is the likelihood ratio in Bayes' Theorem that permits the calculation of posterior probabilities. The posterior (post-identification) probability that the identified suspect is in fact the culprit depends critically on the diagnosticity ratio. The exact calculation of the posterior probability of guilt requires knowledge of the base rate (prior probability that the lineup contains the culprit). However, the higher of any two diagnosticity ratios always produces the higher posterior probability of guilt for all values of the base rate between 0% and 100%. Later in this article, we present posterior probability curves to illustrate this point.

Relative and Absolute Judgments: The Theory Underlying Lineup Format Effects

A traditional lineup allows the witness to make a side-by-side comparison of simultaneously displayed photos or live lineup members, and the witness may choose the lineup member who most closely resembles the offender *relative* to the others. The original theory behind the sequential lineup was that the witness should instead be forced to make an *absolute* judgment about each lineup member (by comparing each lineup member to a decision criterion). Absolute judgments were expected to produce better accuracy than relative judgments (Wells, 1984; Lindsay & Wells, 1985). At the theoretical level, a criticism of this position is that what is meant by *absolute* and *relative* judgments is not totally clear; furthermore, there has been no direct evidence that absolute judgments produce better identification accuracy than relative judgments. Recently, computational modeling has offered some clarification: lineup identification tests of various versions of relative against absolute judgment have revealed that absolute processing produces better overall accuracy (Clark, Erickson, & Breneman, in press). Hence, the theoretical foundation has been placed on a more solid footing. Nevertheless, the superiority of absolute over relative judgments does not necessarily mean that sequential lineups are superior to simultaneous lineups. Likewise, a superiority of sequential lineups over simultaneous lineups does not automatically mean that the superiority is due to greater reliance on absolute judgments. With this caveat in mind, we focus on the question of what the overall literature shows about the sequential-superiority hypothesis.

Operational Specifics of the Sequential Lineup Procedure

At the core of the sequential lineup is the one-at-a-time presentation format. However, the standard sequential lineup is comprised of a package of procedural components that accompany and facilitate this lineup format. For example, the eyewitness does not know how many photos will be shown (a "back-loaded" lineup) nor is able to compare photos side-by-side, each photo requires a decision from the witness before the next is shown, and the lineup is not repeated (no "laps"). As such, simultaneous and sequential lineups differ in more than just the one feature of lineup format. Related to this, a second theoretical criticism leveled at the sequential lineup involves a demand for a more precise accounting of the cause-and-effect relationship between each component of lineup procedure and eyewitness performance. Uncertainty about the precise role of each element in the sequential protocol has yielded a belief for some that knowledge about the sequential lineup is underdeveloped and a poor basis for public policy. However, it is also apparent that the experimental parsing of these lineup features may not be a viable endeavor (for a discussion of this issue, see Lindsay, Mansour, Beaudry, Leach, & Bertrand, 2009a,b; Malpass, Tredoux, & McQuiston-Surrett 2009a,b).

Our objective for this meta-analysis is not to dissect the sequential procedure for theoretical purposes but to more clearly articulate the components of the sequential lineup and thereby to define the parameters for effective sequential practice. Over the years the prescriptive sequential protocol has been differently interpreted or intentionally adjusted by researchers. One such instance is the "stopping rule" employed by some researchers that ends the lineup at an identification, presumably to prohibit the witness from a subsequent decision change. The original sequential procedure (Lindsay & Wells, 1985) required that the witness view each of the photos even if an identification was made early in the lineup. Perhaps not surprisingly, practical questions then arise about the correct or most effective sequential protocol. Related to the stopping rule, Clark and Davey (2005) report a sequential advantage moderated by an order effect: culprit identifications were lower when a "next-best" filler was presented prior to the culprit. Memon and Gabbert (2003a) likewise found that the individual viewing a sequential lineup sometimes spent his or her choice on a similar-looking filler that preceded the culprit. These outcomes have implications for effective lineup composition and procedure; therefore, these aspects of lineup procedure will be addressed in our analysis.

Robustness of Sequential Lineup Effects

A primary benefit of meta-analysis is the examination of patterns across studies, in this case to identify consistent effects of lineup presentation on eyewitness decisions. In addition, moderator variable analyses can reveal the boundaries of lineup format effects. Lineup format impact may be accentuated or attenuated by moderators such as methodological features of the original studies (e.g., child versus adult participants, stimulus materials, lineup structure) or procedural aspects of the lineup (e.g., witness instructions, back-loading, stopping rule).

The robustness of what was dubbed the *sequential-superiority effect* in 2001 has been directly challenged by McQuiston-Surrett, Malpass, and Tredoux (2006), who express concern that the R.C.L. Lindsay laboratory accounted for much of the data to 2001, a potential problem if an unrecognized idiosyncratic factor of the Lindsay lab contributed to the effects obtained. At base, this is an argument about the generalizability of the sequential-superiority effect. McQuiston et al. state: “Bluntly put, outside of the corpus of published studies emanating from the single laboratory, there is no evidence that SEQLs [sequential lineups] are in overall terms superior to SIMLs [simultaneous lineups].” (p. 141). The current meta-analysis will address this claim regarding a Lindsay lab moderator variable, as well as McQuiston-Surrett et al.’s assertion that failure to counterbalance lineup member positions (a point of research design) may undermine a sequential-superiority effect.

An intriguing moderator recently has been posited by Clark, Howell, and Davey (2008) who compared the effects of simultaneous and sequential formats by examining past studies, some that directly compared the two formats and others from a larger corpus of simultaneous lineup conditions. These authors suggest that the frequently biased lineups used in simultaneous-sequential comparisons are unrepresentative of the full range of lineups and that when lineups are unbiased, the sequential advantage may not hold. Thus, a sequential lineup advantage may be predominantly a phenomenon of biased lineups, a moderator that will be assessed through this meta-analysis.

Objectives for the Current Meta-Analysis

The 2001 (Steblay et al.) meta-analysis and the studies upon which the analyses were based have served as a foundation for policy decisions, legal reasoning, and scientific analysis (see e.g., Wells, 2006). Since 2001, the number of experimental investigations of sequential versus simultaneous lineups has accelerated, and the number of independent labs contributing to the dataset has increased dramatically from seven to 23. The central objective of this meta-analysis is to assess the status of the lineup format effects across extant studies and to identify factors that moderate the effects. The new meta-analysis will statistically summarize the substantial new research, integrate it with earlier data, and address procedural and policy questions.

Method

Dataset

Following an electronic search, a review of conference programs, and personal email requests to approximately 50 eyewitness researchers, the resulting papers were screened according to the following criteria: (1) the study provided a statistical test that compared a sequential to a simultaneous lineup format, (2) the statistics required to directly compute a *z*-test and *r* (effect size) were available either within the article or from the author, and (3) the test was for event memory (not a facial recognition paradigm). Although great variability in design and method is acceptable and even desirable as the basis for moderator analyses, we excluded tests of within-subject comparisons of the lineup format manipulation, tests in which multiple culprits were positioned in a single lineup, and those in which witnesses were allowed multiple laps through the lineup (although we allowed first-lap data when participants were unaware of a second-lap option). All tests involved single-culprit (suspect) lineups in which the witness had only one viewing of a lineup (either simultaneous or sequential) for a given culprit.[1] The original data were collected in Canada, the United Kingdom, South Africa, Germany, and the United

States. Information beyond the written report was secured through follow-up contact with researchers.

Forty-nine papers with 72 (non-independent) tests of sequential versus simultaneous lineup format from 23 different labs were found acceptable based on the above criteria. This set includes 31 tests from studies represented in the 2001 (Steblay et al.) meta-analysis plus 28 new papers that offer 41 new tests. The dataset includes work from 1985 to 2010, representing 13,143 witness-participants, with 55 published (76%) and 17 unpublished tests.

Authors who reported participant samples in generic plural terms such as "undergraduate students" or "community residents" were assumed to have included both male and female participants; under this assumption 100% of the studies included both genders. Sample sizes ranged from 32 to 2529, with a mean of 182.54. (One sample of 2529 is an outlier; with this test removed, sample sizes ranged from 32 to 619, with a mean of 150.) All tests employed photo lineups and 89% of lineups were of size six. The culprit was male in 90% of the stimulus materials. Ninety-eight percent of the tests reported culprit exposure times of less than five minutes.

Procedure and Statistics

Two authors of the current paper independently reviewed each article, coded moderator variable information, and calculated decision frequencies. Following Rosenthal (1991), the primary statistics computed were Z as an unweighted test for differences between groups and the correlation coefficient r as an unweighted index of effect size. We have elected to follow the Steblay et al. (2001) meta-analysis and the McQuiston-Surrett, et al. (2006) paper in using unweighted values. The mean effect size for a group of hypothesis tests is referred to in subsequent discussion simply as r. A useful aspect of r in the comparison of eyewitness lineup decisions is that r closely approximates the difference in percentages between conditions. For example, an effect size of $r = .05$ to describe the difference in culprit identifications between the simultaneous and sequential lineup conditions will allow one to correctly surmise that the difference between the groups is approximately five percent.

A meta-analytic *Z* (*Zma*) was calculated by combining *Z*-scores of individual tests of the hypothesis using the Stouffer method (Rosenthal, 1991). This method produces an overall probability level associated with the observed pattern of results. Rather than using 0.00 ("non significant") or 1.65 ("significant") as an estimate for imprecisely reported *Z* values, we only included tests for which *r* and *Z* could be calculated. A fail-safe *N* (N_{fs}) indicates the number of fugitive tests with null results that would be necessary to overturn a significant outcome. Alpha is set at .05 and all confidence intervals are calculated at 95%.

Results

The Steblay et al., (2001) meta-analysis used the phrases *target-present* and *target-absent* to denote whether or not the culprit of the crime was in the lineup. To avoid some confusion that these terms have produced, this meta-analysis substitutes the phrases *culprit-present* and *culprit-absent.* The signed value of *r* and *z* statistics indicates the direction of the obtained results. Positive *r* and *Zma* values denote support of a sequential advantage, that is, eyewitnesses performed better in the sequential lineup condition. Negative *r* and *z* values indicate the opposite: witnesses in the simultaneous lineup condition were more accurate than participants in the sequential lineup condition. See Appendix A for a list of tests.

All Tests

The first analysis entails all 72 tests, each comparing the performance of eyewitnesses between simultaneous and sequential lineup formats. This set of studies includes many tests in which the very best principles of lineup practice were employed, but also some tests in which (usually intentionally) aspects of poor lineup practice were explored. Magnified effect sizes are often associated with specific experimental manipulations such as lineups biased with respect to foil similarity, instruction, or clothing (see, e.g., Blank & Krahe, 2000, Lindsay, Lea, Nosworthy, et al., 1991). A variety of lineup practices in the laboratory helps to mimic real-world variability and to define parameters of lineup format effects.

Eyewitness decisions: Culprit-present lineups. Three outcomes are possible for an eyewitness who views a culprit-present lineup: correct identification of the culprit, an incorrect choice of a filler (a known error) or an incorrect rejection of the lineup (no-choice). Culprit identifications from culprit-present lineups are significantly more frequent with the simultaneous lineup, *Zma* = -9.57, $p < .0001$, $k = 58$, and $r = -.14$, $N_{fs} = 1905$, with a 14% performance advantage (simultaneous lineup: $M = .52$, $CI_{.95}$ [.47, .57]; sequential lineup: $M = .38$, $CI_{.95}$ [.33, .43]; Table 1).[2] Filler pick rate is equal between lineup formats, at 24%; significant choosing rate differences between simultaneous and sequential lineups are therefore represented in the culprit identification rates.

Eyewitness decisions: Culprit-absent lineups. An eyewitness who views a culprit-absent lineup will produce one of two outcomes: correct rejection of the lineup (no pick) or a mistaken identification. Correct rejections are 21% higher with the sequential lineup compared to the simultaneous lineup, *Zma* = 16.45, $p < .0001$, $k = 64$, $r = .22$, $N_{fs} = 6339$, (sequential M = .64, $CI_{.95}$ [.58, .70]; simultaneous $M = .43$, $CI_{.95}$ [.37, .49]). For this dichotomous accuracy measure, mistaken identifications (and choosing rate) in the two conditions are the reciprocal percentages: 36% and 57%.

Eyewitness decisions: The designated innocent suspect. Twenty-seven research tests explored eyewitness reaction to an innocent but similar-appearing suspect planted in a culprit-absent lineup. False identification of the designated innocent suspect was significantly more frequent from the simultaneous lineup (M = .25, $CI_{.95}$ [.18, .32]) than from the sequential lineup (M = .13, $CI_{.95}$ [.09, .17]), *Zma* = 7.95, $p < .0001$, $r = .14$, $N_{fs} = 604$.

Summary. The results for 72 tests of sequential versus simultaneous lineups are remarkably similar to those obtained in the 2001 (Steblay et al.) meta-analysis and provide evidence for reliability of the obtained effects (Table 1).[3] When the culprit is present in the lineup, witnesses in the simultaneous lineup condition make significantly more culprit identifications. When the culprit is not in the lineup, participants in the sequential lineup condition make significantly fewer mistaken identifications. The stem-and-leaf plots for effect sizes of culprit present and absent conditions (Figures 1 and 2) illustrate the distribution of effect sizes around the means of -.14 for culprit-present lineups and +.22 for culprit-absent lineups.

Moderator variables. Subsequent analyses explore moderator variables, to describe the conditions under which differences in eyewitness accuracy between sequential and simultaneous lineups (effect sizes) become larger or smaller. It is important to first consider the uses of moderator analyses and the limits of their informational value. The key variable for comparison in this meta-analysis is lineup format: within each of the 72 tests was a direct comparison between the two conditions of the independent variable (a sequential and a simultaneous lineup). The experimental design of each study allows us to draw cause and effect conclusions about the impact of lineup format on eyewitness decisions both within each original study and in the aggregate results of the meta-analysis.

Moderator variable analyses offer a different and more limited form of evidence. A moderator analysis compares *between* studies a variable that was not experimentally manipulated *within* each study; a group of studies that possess some characteristic is compared to another group of studies that do not possess that characteristic. Moderator analyses can offer direct evidence about some types of descriptive claims. For example, a conjecture that unpublished studies do not demonstrate the sequential advantage that is present in published work can be investigated by separating published from unpublished studies (publication status as the moderator variable) and examining whether the sequential advantage is absent in the unpublished work as predicted. But, because this comparison is pseudo-experimental, no evidence about cause-and-effect is established. (We will not be able to claim that the unpublished status of a group of studies *caused* the effect size to be different from published studies.) We cannot know *why* a specific effect occurred; only that it did. Without a true experimental design, confounding variables prohibit causal conclusions.

Moderator Variable Analysis: Robustness of Lineup Effects (Table 2)

Eyewitness age. The 2001 meta-analysis (Steblay et al.) indicated that a sequential lineup was not of benefit when eyewitnesses were children. More recent research (e.g., Memon & Gabbert, 2003a) indicates that older adults have difficulties with lineup decisions, generating considerable mistaken identifications. Table 2 illustrates that tests with older adults and those with children show a significant advantage of the simultaneous lineup in a culprit-present lineup condition, and older adults demonstrate a significant sequential lineup advantage in a culprit-absent condition. However, the general lesson from the few available tests is that older witnesses and children make large percentages of errors. When the culprit is absent from the array, the sequential format seems to inhibit choosing somewhat with older adults, but overall these tests show relatively high filler pick rates with both formats: 50% with sequential and 74% with simultaneous lineups. Similarly, in these tests, children chose from both sequential and simultaneous lineups at high rates (84% when the culprit is present, 75% when the culprit is absent), with very high error levels regardless of lineup format. In short, children and older adults show significantly different (and profoundly poorer) eyewitness performance compared to the (non-older) adult population regardless of whether they are using the simultaneous or the sequential format. Our subsequent analyses exclude tests of older adults and children.

Lindsay lab. Recent years have seen sequential lineup testing move well beyond the lab of its originators (Lindsay & Wells, 1985). Tests from the "Lindsay lab" (defined as any study for which R.C.L. Lindsay is a co-author or that originated from his Queens University lab) now comprise 36% of the adult witness dataset and 17% of the full-design dataset described below. This allows the opportunity to explore the robustness of lineup format effects across laboratories. Results of the comparison (adult witnesses only) reveal that the significant effects for both culprit-present and culprit-absent lineups produced through Lindsay's lab are reliably evident in other laboratories and vice versa (Table 2).[4] One difference is that the Lindsay lab generates a significantly larger sequential advantage in the circumstance when a designated innocent suspect is planted in the culprit-absent lineup (false identification reduction of 20% in the Lindsay lab vs. 5% in other labs), $t(23) = 4.37$, $p < .001$. Lindsay, Lea, and Fulford (1991) explain that false identification rates should be inflated to the extent that an innocent suspect is physically similar to the offender, and the Lindsay lab has demonstrated a strong sequential advantage in this situation. However, physical similarity is only one means through which an innocent suspect may end up in a lineup. Importantly, tests from other labs without a designated innocent suspect produce the sequential advantage with the broader measure of mistaken identifications of any filler in a culprit-absent lineup. Significant lineup format effects are not exclusive to the Lindsay lab.

Publication status. Both published and unpublished tests show the common pattern of simultaneous benefit when the culprit is present and sequential benefit when the culprit is absent.

There is a significant difference between published and unpublished studies in the size of effect for the culprit-present condition, with a smaller effect for published work, $t\ (46) = 1.93$, $p = .03$.[5] One can speculate that sound methodological reasons may have kept certain tests out of peer-reviewed journals. With some exceptions (e.g., newer unpublished studies may ultimately move into scientific journals), unpublished studies may include small sample size, unrefined pilot projects, lack of experimental controls, or methodological details long forgotten and thus unavailable for peer review. These shortcomings also may contribute to the experimental effects obtained in unpublished work.

All tests, published and unpublished, typically are included in meta-analytic calculations in order to work with an increased amount of information, as calculated above. At the same time, an argument can be made for examining only published work as a means to meet Daubert (1993) criteria (see e.g., Deffenbacher, Bornstein, Penrod, & McGorty, 2004), and the following diagnostic analyses are useful for that purpose.

The Full Diagnostic Design Dataset ("Full Design"): The 2 X 2 Gold Standard

Thirteen labs (27 published tests) used, at a minimum, a full 2 X 2 fully-randomized factorial design to explore lineup format effects (sequential/simultaneous lineup format X culprit present/absent lineups) with adult eyewitnesses. We refer to this subset of studies as the full diagnostic design dataset because the fully randomized design within each study allows us to draw cause-and-effect conclusions about the impact on eyewitness decisions of lineup format between comparable culprit present and absent conditions. The full diagnostic design dataset is the "gold standard" because the independent variables are totally un-confounded with study differences (e.g., the view that witnesses had, the similarity of the fillers); any study differences can contribute noise within this 2 X 2 design, but study characteristics are not confounded.

Importantly, this *full diagnostic design dataset* also allows us to protect subsequent diagnosticity calculations from the influence of uneven research design and to determine diagnosticity ratios with the strong scientific rigor of published work (see Appendix A). For brevity, we refer more simply below to the *full design* dataset.

One additional criterion was deemed necessary for inclusion of a test in the full design dataset: that the performance of witnesses must be clearly above chance. There are obvious reasons why meta-analytic researchers are concerned with excluding studies that have methodological characteristics that do not provide a proper test of the hypothesis under consideration. In the current work, for example, studies that provide such poor views of the perpetrator that witnesses could not be expected to perform regardless of whether the test was simultaneous or sequential ought to be excluded because they do not provide an opportunity for the simultaneous to show its advantage in hits or the sequential to show its advantage in correct rejections. Unfortunately, researchers commonly do not provide enough information to make these types of judgments (see McQuiston-Surrett, et al., 2006, for a similar point). Furthermore, even in cases where relevant information is reported, subjective judgments would have to be made as to whether the concern (e.g., how poor was the view? How poor would it need to be?) should result in exclusion of the study from the full design dataset. Fortunately, there is an objective measure of whether methodological problems prevent a proper test of the simultaneous versus sequential question. Specifically, every study can be examined for whether the performance of the witnesses was appreciably above chance. Using a criterion that requires witness performance to be above chance takes care of numerous problems that could be hidden in the methods and provides a more objective criterion for inclusion and exclusion. For instance, if the view was poor enough for exclusion, then witness performance should not be above chance. Similarly, studies that used fillers who are near-clones of the culprit, used a poor photo of the culprit, used a highly biased lineup in which the innocent suspect stood out, or used a near-clone of the culprit as an innocent suspect replacement in the culprit-absent lineup would tend to yield chance performance. In fact, the reason to be concerned about these problems (e.g., poor view, use of a near clone of the culprit in the absent lineup) is precisely because they can yield chance performance. Hence, rather than making subjective judgments about these characteristics of studies in order to decide whether they should be included in the full design dataset, we used objective procedures for deciding whether performance was appreciably above chance.

Two common metrics can establish whether eyewitness identification is above chance for a given set of data: (1) the relative choice rates for the culprit in the culprit-present lineup versus an innocent person in the culprit-absent lineup and (2) the relative rates of correct rejections and false rejections (Wells and Penrod, in press). We chose a liberal criterion for inclusion of a test in the full design dataset. Only one of these two metrics had to meet or exceed 10% in either the simultaneous or the sequential dataset in order for a study to be included. In other words, any given study that included the full 2 X 2 test had four opportunities to show that just one comparison was above chance level: (1) did culprit identifications exceed innocent suspect identifications for the simultaneous lineup? or (2) did culprit identifications exceed innocent suspect identifications for the sequential lineup? or (3) did correct rejections exceed false rejections for the simultaneous lineup? or (4) did correct rejections exceed false rejections for the simultaneous lineup? If any one of these four were above the 10% criterion, the study was included. Using this criterion, three tests from three different research teams were excluded from the full design dataset: Douglass & McQuiston-Surrett (2006), Gronlund, et al (2009), and Steblay (2010).

To illustrate the problem of “at chance” witness performance for a comparative test of sequential versus simultaneous lineup performance, we focus on one large study that failed all four metric tests, namely Gronlund et al. (2009). In the Gronlund, et al. dataset, the correct rejection rate did not exceed the false rejection rate by 10% for either the simultaneous or the sequential lineup. Nor did correct identifications of the culprit exceed false identifications of the innocent replacement by 10% in either the simultaneous or the sequential condition. In fact, among those making identifications from the simultaneous lineup, 65% in the culprit-present lineup condition identified the culprit and 63% in the culprit-absent lineup condition identified his innocent replacement. Similarly for the sequential lineup, 60% identified the culprit from a culprit-present lineup and 59% identified his innocent replacement from a culprit-absent lineup. The pattern overall indicates that the innocent replacement was a near clone of the culprit. This is perhaps not surprising given that the innocent replacements (two were used) were found by searching the Florida Supervised Offenders database, a repository of thousands of photos, (http://www.dc.state.fl.us/activeoffenders/search.asp), from which 28 were selected that "we judged looked like the perpetrator" (p. 143). From the 28, two final photos were selected from the three that scored highest in rated similarity between each face and the culprit (from 80 raters). Given that the result was near-chance eyewitness lineup performance, it is not surprising that Gronlund et al. found no overall advantage for sequential or simultaneous lineups. More importantly, this was not a reasonable or informative test of sequential versus simultaneous lineups.

Overall results. The pattern of results with the full design dataset (Table 3) is very similar to that obtained with the 72-test dataset (Table 1).

Choosing rates. The simultaneous lineup produces significantly higher pick rates than does the sequential lineup in both culprit-present and culprit-absent lineups, leading to more culprit identifications when the offender is in the lineup and more mistaken identifications when he is absent. The 24 tests that provide choosing rates for culprit-present lineups indicate an average sequential lineup choosing rate of .61, $CI_{.95}$ [.54, .68] and an average simultaneous lineup choosing rate of .76, $CI_{.95}$ [.72, .80].

Eyewitness decisions: Culprit-present lineups. The significant advantage of the simultaneous lineup for correct identifications in the culprit-present lineup condition is trimmed by six percentage points in this full design dataset compared to the full 72-test set, to an average eight percent, with the range of effect sizes from -.32 to +.21 (r = -.08). Confidence intervals overlap somewhat between sequential lineup *M* = .44, $CI_{.95}$ [.37, .51] and simultaneous lineups *M* = .52, $CI_{.95}$ [.47, .57].

Eyewitness decisions: Culprit-absent lineups. The significant sequential lineup advantage for reduced mistaken identifications within the culprit-absent lineup condition remains virtually the same in this subset of tests (r = .23). Moreover, all but one of 27 effect sizes for culprit-absent lineups are positive, ranging from -.04 to .78—a robust sequential advantage (Figure 2). Mistaken identifications are significantly more frequent from a simultaneous than a sequential lineup; simultaneous *M* = .54, $CI_{.95}$ [.47, .61]; sequential *M* = .32, $CI_{.95}$ [.25, .39]. Mistaken identification of an innocent suspect planted in the lineup is significantly more frequent from a simultaneous than a sequential lineup (simultaneous M = .28, $CI_{.95}$ [.19, .37]; sequential M = .15, $CI_{.95}$ [.07, .23]).

Diagnosticity. The sequential lineup produces a diagnosticity ratio of 7.72, the simultaneous lineup a ratio of 5.78.[6] Identification of the suspect from a sequential lineup is

1.34 times more diagnostic compared to an identification from a simultaneous lineup. If the rate of identifying the known-innocent suspect in the culprit-absent condition is used as the ratio denominator, this diagnosticity index yields 1.86 for the simultaneous lineup and 2.94 for the sequential lineup. The sequential lineup is 1.58 times more diagnostic.

Remaining (Non-Diagnostic) Data

The complement to the full design dataset is the 45 remaining tests, consisting of unpublished studies, those that involve very young or older witnesses, those that do not include both culprit-present and absent lineup conditions, and those that do not meet the criterion for testing above chance levels of identification. In this set of tests (Table 2) the typical sequential advantage is present in culprit-absent lineups, however, the culprit-present lineup effect size is significantly larger compared to the full design set (-.19 vs. -.08), $t\ (56) = 3.08$, $p = .003$, 2-tailed. This outcome tells us that a reduction of diagnosticity for the sequential lineup is associated with factors of study sample, design, and quality. The comparative diagnosticity of sequential (5.07) and simultaneous (5.31) lineups is very close in this dataset.

Moderator Variable Analysis: The Full Diagnostic Design Dataset and System Variables[7]

The intent of this section is to address system variables, controllable aspects of identification procedures that may affect eyewitness performance (Wells, 1978). Two core system variables—lineup size and use of a cautionary instruction that the culprit “may or may not be in the lineup” (Steblay, 1997)—were implemented almost uniformly across the tests of the full design dataset and thus do not offer meaningful moderator analysis.

Lineup construction method. Two primary approaches to lineup construction have been employed in the lab. *Match-to-description* is considered by most researchers to be a superior method of constructing a fair lineup (Luus & Wells, 1991), the lineup fillers based on the eyewitness's description of the culprit. A second lineup formation strategy involves a *match-to-culprit* determination, in which lineup fillers for both culprit-present and culprit-absent lineups are based on the culprit's appearance. The comparative results between the two techniques indicate that both *match-to-description* and *match-to-culprit* outcomes align with the common pattern of format differences: the simultaneous lineup produces significantly more culprit identifications from culprit-present lineups and the sequential lineup significantly reduces mistaken identifications from culprit-absent lineups (Table 4). Notably, a match-to-description method is associated with higher diagnosticity under both lineup formats, but in conjunction with a sequential format shows the highest diagnosticity (10.00) obtained for the system variables detailed in Table 4.

Back-loading. Information about back-loading was available from 26 tests. In 22 of the 26, the sequential lineup was back-loaded, accomplished by hiding from the witness the exact number of photos to be shown. In the four tests in which no back-loading was employed, the simultaneous lineup advantage in the culprit-present condition is absent—the sequential lineup culprit identification rate (58%) matched that of the simultaneous lineup (57%). However, the four effect sizes range from -.09 to .07, suggestive of unaccounted for variability.

Description of the culprit prior to lineup. In the field, witnesses are often asked to provide a description of the culprit prior to the lineup, a task duplicated in the experimental protocols of some labs. This procedure may lead to some reduction in eyewitness accuracy, a phenomenon referred to as *verbal overshadowing* (Meissner & Brigham, 2001). In 14 tests for which witnesses were required to describe the culprit before the lineup, the simultaneous advantage in culprit-present lineups is (non-significantly) smaller than for tests in which a description of the culprit was not required.

Stopping rule. The witness who views a simultaneous lineup is allowed to compare photos before deciding on any one (or none) of them, and a first inclination to identify a particular photo may be stifled if another lineup member is found to be a closer match to memory. Conversely, in the sequential lineup, it is intended that the witness make a decision one photo at a time, and researchers who employ a "stopping rule" take a first "yes" decision as final and show the witness no additional photos. For most research teams (77%), the lineup photos were continued after an identification, and, if instructions did not forbid it, a witness could make a second identification or change a decision. How this witness's decision is recorded—a determination must be made as to which of multiple responses is the true decision of that eyewitness—is a protocol consideration.

A significant sequential advantage in culprit-absent lineups is apparent no matter what the stopping policy or which decision governs the handling of multiple identifications. Also, the diagnostic benefit of the sequential lineup surpasses the simultaneous lineup under any of these strategies. In the culprit-present condition, however, the stopping rule moderates the size of eyewitness performance differences. Although the simultaneous advantage remains statistically significant, the difference in culprit identifications between lineup formats grows to 17% for "stopping," and shrinks to 5% for the "continue-to- the-end" studies, a significant difference in effect sizes, $t(24) = 2.05$, $p = .05$. It is puzzling, however, that sequential lineup culprit identification rates are the same across the two groups of studies, at 45%, that is, whether a stopping rule is used or not. On the other hand, the stopping rule is associated with a *simultaneous* lineup jump from 50% culprit identifications (in studies in which the sequential lineup continues to the end) to 62% (in studies in which the stopping rule is used). We address this confusing outcome more fully in our discussion.

The decision as to how to count multiple choices by a witness—using a first decision rule or as an immediate error—should make a difference when the culprit is present in the lineup. There is a lack of stability associated with the first-response rule, effect sizes ranging from -.26 to .21; the outcome of this rule for a given witness's performance presumably depends on whether the witness's first pick is the culprit or a filler. Sequential (vs. simultaneous) lineup performance should be more affected by a rule that counts multiple picks as errors, particularly under a circumstance in which an early pick of a filler is corrected (and recognized by the witness as an error) when the culprit appears later in the lineup. In this scenario, one would expect a simultaneous advantage, as is the case. A record of eyewitness verbal responses is not typically a part of laboratory protocol. Lab researchers must concede that a "1st-choice," "last choice," or "immediate error" laboratory rule may unintentionally blur the record of eyewitness accuracy in the sequential condition.

Moderator Analyses: Laboratory Method (Table 4)

Position/order effects. The common pattern of lineup effects remains evident whether the researcher held the culprit/suspect in a stable position (at position 3 or beyond) or used a full rotation of the offender through the lineup (avoiding position 1), and whether or not counterbalancing is used. There are non-significantly smaller effect sizes for tests that used partial rotation (in culprit-present and designated innocent suspect conditions), and a larger effect size (.19) for two tests that did not counterbalance innocent suspect position (*r*s = .08 and .30). The lack of statistically significant differences in these conditions (despite seemingly large differences in effect sizes) is likely due to the variability in outcomes across a small number of tests; for the same reasons, it is difficult to draw firm conclusions about these effect size differences.

Control of experimenter effects. Experimenter effects in lineup research are controlled through a variety of strategies. To limit unintentional cues, the experimenter may leave the room during the lineup, stand behind or away from the participant, not directly handle the lineup photos, allow the witness his or her own pace through the photos, engage in only scripted verbal exchange, and/or use a computer to present the lineup. Some studies employ the combination of a group setting and private (paper) witness response as a means to minimize the administrator's interaction with any one participant. Most researchers (22 tests) in the full design dataset report use of multiple strategies to limit experimenter effects, and their results follow the prevailing pattern of eyewitness performance differences between lineup formats (Table 4). Across the entire 72-tests, there are more tests that report controls for experimenter effects in the full design dataset (78%) than in the remaining tests (40%), $Z = 3.17$, $p < .01$. Reported control for experimenter effects also is one aspect of research design that is more common in published work (60%) than in unpublished work (35%), $Z = 1.81$, $p < .05$.

Presence of a designated suspect in the culprit-absent lineup. A designated innocent suspect is a filler that most closely matches the description of a culprit or is rated as physically most similar to the culprit. When placed in a laboratory culprit-absent lineup, this person represents the worst-case scenario that may unintentionally occur in field practice. The suspect is similar in appearance to the culprit and thereby can be expected to draw a disproportionate number of witness picks. Indeed, in this dataset, the innocent suspect draws 39% of the picks from a sequential lineup and 49% of the picks from a simultaneous lineup, well above the rate expected by chance in a fair lineup (16.6%), Zs > 1.65, ps < .05. In this respect the lineup is biased against the innocent suspect. A comparison between lineups constructed with a designated innocent suspect (*biased* against a suspect) and lineups that do not feature a specific innocent suspect (*not biased* against a specific suspect) indicate no significant difference in effect sizes for the culprit-absent condition. The sequential advantage is not moderated by this type of lineup bias.

Lineup fairness. Table 4 lists procedures used to arrive at a level of fairness in lineup structure (selection of fillers) deemed appropriate by the researcher, although these procedures do not directly translate to greater or lesser fairness. The common significant pattern of simultaneous advantage in culprit-present lineups and sequential advantage in culprit-absent lineups is apparent for tests in which lineup fairness has been determined using mock witness procedures and/or reported with a fairness index (arguably *unbiased* lineups), but not for lineups developed through a ranking of visual similarity.

Additional moderators. Stimulus mode (live, video), exposure to culprit (< 10 s, 10-20 s, 60-75 seconds), and delay (< 30 minutes, > 24 hours) were also examined and produced no significant moderator impact.

Regression Analysis: All Adult Data (60 tests)

The moderator analysis display on Table 4 indicates that the common pattern of simultaneous lineup advantage for correct identifications in culprit-present lineups and sequential lineup advantage for reduced mistaken identifications in culprit-absent lineups runs through almost all tested variations in lineup procedure. This indicates a robust phenomenon. It is also true that these factors are confounded within studies. In an attempt to untangle the effects of procedural and methodological components in sequential lineup performance, we moved to a regression analysis.

The full dataset (adult witnesses) was employed to conduct regression analyses using predictors of stopping rule, verbal overshadowing, back-loading, lineup construction method, experimenter expectancy control, and target position. These variables did not initially produce a statistically significant model for prediction of effect size in culprit-present lineups. Not unexpectedly, high collinearity among the variables is a problem. Variables with eigen values close to zero were removed, and a subsequent analysis resulted in a statistically significant predictive model. Eigen values indicated only a single variable—stopping rule—that contributed significantly to the outcome. A subsequent stepwise regression also indicated stopping rule to be a significant predictor of culprit-present effect size (t = 3.79, p = .001, B = .504) as was back-loading, (t= 3.10, p = .004, B = .413). Collinearity remained an issue, however, making outcomes tentative at best. Eigen values signaled that back-loading remained a minimal (and highly correlated) contributor to variance. And, as will be discussed below, unknown contributors to a stopping rule moderator effect in culprit-present lineups severely limits the interpretation of this analysis. Effect size for culprit-absent lineups was also tested as a dependent measure; no effective predictive model emerged.

The primary contribution of this new meta-analysis is in its description of sequential and simultaneous lineup effects on eyewitness accuracy derived from the full design dataset. Subsequent discussion addresses considerations of how these results can be viewed from legal and public policy perspectives.

Discussion of Findings

We first discuss key outcomes of the meta-analysis, including limitations of the analyses. Then, we discuss numerous policy considerations. Clearly, the results have implications for matters of policy, but conversely, there are many policy considerations that can put a very different light on the findings.

A primary finding is that the full 72-test dataset from 23 different labs involving 13,143 participant-witnesses yields overall results that are highly similar to those reported in the 2001 (Steblay et al.) meta-analysis. The sequential lineup reduces mistaken identifications from culprit-absent lineups. The simultaneous lineup produces more culprit identifications when the offender is in the lineup. Furthermore, the data do not support the contention that an individual lab is driving this pattern of results.

Also of vital importance is the fact that there is now a substantial number of published studies (27) from numerous labs (13) that used the full 2 (simultaneous or sequential) X 2 (culprit present or absent) design. These fully randomized studies represent the only database that can reasonably support cause-and-effect claims about the comparative advantage between lineup formats. The full diagnostic design dataset is the "gold standard," and it reveals that the difference between sequential and simultaneous lineups in rates of culprit identification is 8% (in favor of the simultaneous lineup) compared to the 15% difference in the 2001 meta-analysis (that did not exclusively use 2 X 2 designs in the analyses). The full design dataset produces a diagnosticity ratio that is higher for the sequential (7.72) than for the simultaneous (5.78) lineup: this is the *sequential-superiority effect*. The full design dataset also reveals higher sequential lineup diagnosticity in the circumstance when a similar-appearing (to the culprit) innocent suspect is in the lineup (sequential 2.94; simultaneous 1.86).

A Special Caution about Moderator Analyses

We searched for a number of moderators of the sequential-superiority effect that have been suggested by various researchers, but our search resulted in few significant outcomes. For example, no significant moderation was evident for stimulus mode (live, video), exposure to culprit (< 10 s, 10-20 s, 60-75 seconds), delay (< 30 minutes, > 24 hours), lineup bias associated with a designated innocent suspect, counterbalancing strategy (or not) for position of suspect, sequential lineup back-loading, and whether or not the witness gave a description prior to the lineup. Lineup effects were moderated by study sample (young children and the elderly), by publication status, and in studies using the stopping rule.

We urge special caution in the interpretation of study-factor moderators. In fact, one finding strongly signals the potentially-misleading nature of moderator analyses, specifically, the significant but strange "effect" of the stopping rule within culprit-present lineups. Studies employing a "stopping rule" showed a simultaneous lineup jump in accuracy from 50% culprit identifications (in studies in which the sequential lineup continues to the end) to 62% (in studies in which the lineup stops after an identification is made). The simultaneous lineup condition should not be affected by a stopping rule because the stopping rule only applies to a sequential lineup; hence, it was not a stopping rule that produced an increase in culprit identifications. Clearly, studies that used the stopping rule somehow differed from studies that did not, in a way that cannot be fully parsed. This is striking evidence and a reminder that we can confidently interpret a relationship as causal only when direct comparisons are tested within the same study (i.e., a direct comparison of stopping rule versus no stopping rule within a study).

In the parlance of meta-analysis, this problem is the result of multi-collinearities that cannot be fully removed from a cross-study analysis because they are confounded within study, often in various combinations and not reported or measured. Outcome differences between one study and another incorporate multiple unrecognized and uneven influences. These may include such factors as differential lighting conditions, photo quality, the extent to which the culprit's photo captures his "normal look", the extent to which the witnessed event captured the full attention of the witnesses, and so on. This brings us back to the reason that the full design dataset is the only firm basis for cause-and-effect interpretations. In the full design dataset, differences between studies merely create noise (not confounds) in estimating the overall sequential-superiority effect.

The Base Rate for Culprit Presence Does Not Affect Sequential Superiority

It is tempting to assume that the superiority of the sequential lineup is dependent on the prior (base rate) probability that the culprit is in the lineup. In fact, however, Bayesian statistics clearly show that a higher diagnosticity ratio for the sequential lineup results in a higher posterior probability that an identified suspect is the culprit; this is true across all possible prior (base rate) probabilities except 0.0 (zero) and 1.0. The prior-by-posterior curves of Figures 3 (based on the full design dataset) and 4 (based on the designated innocent suspect dataset) illustrate this point. In a prior-by-posterior graph, the straight line is the "identity line" and represents the posterior probability if there was no diagnostic value of the tested procedure. Consider the .15 point on the prior (base rate) probability axis (the x axis) for Figure 3. If the witness identifies the suspect using the simultaneous procedure, the posterior probability that the suspect is the culprit rises to a value of .50, whereas if the sequential procedure is used the posterior probability value rises to .59. Or, if the prior probability is .50, the simultaneous lineup produces a posterior of .85 whereas the sequential lineup produces a posterior of .89. Although the diagnosticity of both simultaneous and sequential is lower for the designated innocent suspect dataset, the advantage of the sequential is even greater in this dataset (Figure 4): If the prior probability is .15, the posterior probability for the simultaneous is .25 (versus .34 for the sequential); if the prior is .50, the posterior for the simultaneous is .65 (versus .75 for the sequential).

In both datasets, the maximum difference between simultaneous and sequential lineups occurs when the base rates are lower (maximum sequential advantage is when base rate is .12 for the larger full design dataset and .31 for the designated innocent suspect dataset). The general point, however, is that the advantage (higher posterior probabilities of guilt) favors the sequential lineup across all possible prior (base rate) probabilities between 0.0 and 1.0. That is not a unique feature of simultaneous versus sequential lineups; it is a mathematical imperative that the higher of two diagnosticity ratios yields the higher posterior probability. We display these figures as visual proof to dispel any notion that the advantage of the sequential exists only if the base rate for culprit presence is extremely low.

The Sequential-Superiority Effect is Not Merely the Result of Lower Choosing Rates

It is easy to show that the sequential advantage (in diagnosticity and posterior probability) is not merely the result of lower choosing rates per se. Suppose, for instance, the rate of culprit identifications dropped from 51% with the simultaneous lineup to 31% with the sequential lineup and the rate of culprit-absent filler identifications dropped from 55% with the simultaneous to 35% with the sequential. In this example, the sequential lineup has dropped the choosing rate by 20% for both the culprit and for the absent lineup members. But, in this case, diagnosticity would be higher for the simultaneous than for the sequential lineup. The point is that diagnosticity is greater for the sequential in the current data because the *ratio* of culprit identifications to misidentifications of the innocent suspect is greater with sequential lineups, not because the rate of choosing overall is less. In other words, the sequential-superiority effect observed here depended very much on the fact that the sequential lineup reduced the rate of identifying the culprit by only 8% but reduced choosing in the culprit-absent lineup by 22%.

Locating the Sequential-Superiority Effect

There are multiple possible sources for the sequential-superiority effect, such as the one-at-a-time display of photos, the witness not knowing how many photos are in the lineup, or the requirement that the witness makes a decision about each photo. There is also the fact that a sequential lineup prohibits the witness from fully determining if some characteristic of one photo (e.g., the color of the background) is unique to the set. It is quite possible that it is the combination of these factors that is important. It was not our purpose to tease apart which elements of the sequential procedure contribute to the sequential-superiority effect. Instead, we attempt to articulate the components for an effective sequential procedure. The set of recommendations that makes up the sequential protocol includes requirements for lineup construction, instructions to the eyewitness, format, and procedural features. All tests in this meta-analysis adhered to the rules of a single-suspect lineup with at least four (usually five) fillers, a restriction to a first identification attempt, and a single viewing of the lineup by a witness. The sequential procedure prohibited the witness from side-by-side comparison of lineup members or return to previous photos, and a yes/no decision for each photo was necessary before moving to the next. The tests employed the recommended cautionary instruction to the witness that the true perpetrator may or may not be in the lineup. The modal pattern of eyewitness response—increased culprit identifications from the simultaneous lineup and reduced mistaken identifications from the sequential lineup—was generated from a set of tests that adhered to these requirements. In addition, this pattern is associated with a match-to-description lineup construction method and with tests that employed lineup fairness checks for functional/effective size—thus, arguably unbiased lineups. We now can predict a reliable pattern of eyewitness performance given adherence to these core recommendations.

Discussion of Policy-Related Matters

At this point, it seems unlikely that additional data comparing the two lineup procedures is going to significantly alter the basic pattern of demonstrated eyewitness decisions; the sequential procedure yields fewer suspect identifications, but when obtained, the identifications are more likely to be accurate. Because there is a trade-off between reduction of misidentifications and reduction of culprit identifications, the simultaneous/sequential decision is a policy matter. Science can describe the nature of the trade-off but cannot dictate which is better for practice. In order to better understand the nature of the decision at hand for law enforcement and policy makers, it can be useful to consider a hypothetical and to pose new questions about what these data imply for policy. We offer several pertinent perspectives in the following sections.

The Status-Quo Hypothetical

For more than a century, the status quo has been the simultaneous lineup. Imagine the reverse: suppose that the status quo had been the sequential lineup and that in the 1980's researchers came forward with an alternative called the simultaneous lineup. The current set of meta-analytic data culminates over a 25-year period. Would these current data then lead to a serious call for a switch from sequential to simultaneous lineup procedures, when the change would create a 1.62 greater likelihood of mistaken identification in exchange for a 1.12 greater likelihood of identifying a guilty suspect? Would any law enforcement agency elect a procedure that doubles the risk for identification of an innocent similar-appearing suspect? We think not. And yet, when the status quo is the simultaneous lineup, many seem to consider it risky and unwise to switch to the sequential lineup.

Our hypothetical scenario leads to a somewhat different perspective about the simultaneous versus sequential controversy. This is not surprising given the status quo bias. As a central component of Prospect Theory (Kahneman & Tversky, 1979), the status quo bias is explained as a cognitive partiality that arises through a combination of an endowment effect (assignment of higher value to what one already possesses) and risk aversion. A new proposal is typically evaluated with the status quo as a reference point. Disadvantages of the alternative often loom larger than advantages, and risk aversion prompts a weighting of loss avoidance more heavily than the acquisition of gains (Kahneman, 2003). In that light, the sequential-simultaneous policy decision can be seen as a fascinating example of the unsettling impact of perceived potential losses in correct identifications (culprit-present lineups) even when the offsetting advantage of accuracy gains (culprit-absent lineups) is larger. Without the buttress of a status quo bias, the simultaneous format is less compelling. We will return to consideration of the perceived potential losses of the sequential lineup. First, however, we wish to place lineup performance in the broader perspective of eyewitness identification evidence.

A Reliable Witness Should Be Able to Handle the Sequential Procedure

Eyewitness identification evidence has a profound impact on trial outcomes. An identification decision by an eyewitness can be the primary cause of a person serving a long prison sentence (or even receiving a death sentence). This path has been observed repeatedly in DNA exoneration cases. But these DNA exoneration cases are the "lucky" ones for whom there was DNA evidence powerful enough to trump an eyewitness.

The legal system clearly recognizes the concept of balancing the probative versus prejudicial value of evidence. In other words, the impact of evidence (its prejudicial value) should not exceed its true evidentiary (probative) value. In the context of the incredible weight given to eyewitness identification evidence, there must be a presumption of high probative value, i.e., that eyewitness memory is highly reliable. Thus, we find it curious that critics of the sequential lineup believe that witnesses need to view all the lineup members at once. Why do witnesses need to see what the remaining lineup members look like or know how many more photos will be shown before deciding whether a specific individual is the culprit? Somehow, there is a fear that a witness cannot manage to reject fillers and pick out the culprit if the decisions have to be made sequentially. But consider: If an eyewitness has a good enough memory to have a weighty impact on the fate of a suspect, should not the eyewitness be able to pick that person out from a sequential array? Conversely, it can be argued that an eyewitness who "needs" a simultaneous lineup is a witness whose memory is not strong enough to carry the burden of determining the fate of a suspected person.

A reliable witness should not need to compare a face to the remaining photos. A reliable witness should be able to reject fillers individually. The eyewitness who pauses part way through the sequential procedure and says "can I see the rest before I decide whether this is this one?" is divulging something that only the sequential procedure can reveal; this witness wants to make a mere relative judgment. Should the system permit these comparative judgments, or should the system do what it can to force witnesses to make decisions based more on actual recognition?

Reconsideration of the Order-Effect

Actual recognition is important. Related to this issue is a finding that has led to criticism and consternation about the sequential lineup: the demonstration of an order effect. Specifically, when using the sequential lineup, witnesses sometimes "spend" their identification on a similar-looking filler before they reach the culprit's photo (Clark & Davey, 2005; Memon & Gabbert, 2003b). Among researchers, there tends to an automatic negative reaction to order effects. The researchers know that in their experimental design the actual culprit followed the similar-looking filler and thereby problematically reduced overall "hits." But, is the order effect a problem when one considers the broader perspective of a lineup's evidentiary purpose? Is it not the objective of a lineup to weed out witnesses who are prone to identify a person who is merely similar to the culprit? Yes, this witness might have picked the culprit if the chosen filler had not come up first in the sequential array, but why was the witness willing to *identify* an innocent person who merely possessed features similar to the offender? What does this decision say about the witness? In the real world, the photo that follows the similar-looking filler could be an innocent suspect and the “order effect” saved the innocent suspect from being falsely identified by a witness with a limited memory of the culprit.

From this perspective—and in line with the traditional reason for using a lineup in the first place—the sequential lineup does a better job. Good fillers that precede the suspect in a sequential lineup are lures, filters, or separators of weak witnesses (whose memory is not good enough to reject these fillers) from strong witnesses (who readily reject these fillers). The sequential lineup requires witnesses to reject fillers before (and after) encountering the photo of the suspect in the lineup. The witness who can reject good fillers is a stronger witness; a witness whose decisions can be more trusted. Hence, the sequential lineup is a higher standard.

The Sequential Lineup Spoils Fewer Witnesses for Later Lineups

One of the unsung virtues of the sequential lineup is that the higher correct rejection rate (22% fewer identifications for culprit-absent lineups) "saves" these witnesses for a possible later lineup that includes the actual culprit. Often forgotten is that a witness who identifies a filler is, in effect, a spoiled witness; that witness cannot then view another lineup should the police later find the actual culprit. The implications of this are huge. Out of every 100 culprit-absent lineups shown, 22 fewer witnesses will pick someone if the lineup is done sequentially rather than simultaneously. These are 22 more witnesses (compared to the simultaneous procedure) who could still credibly identify the culprit if the real culprit were shown to them later. With the simultaneous lineup, these 22 witnesses have reduced their credibility by choosing a filler and are thus unlikely to be shown a second lineup.

Previously hidden in all of these analyses, therefore, is the fact that the sequential lineup will pick up some unknown number of additional culprit identifications as an investigation proceeds. In fact, our best estimate is that the 22% additional witnesses "saved" by the use of the sequential procedure when they were shown a culprit-absent lineup could yield another 10% identifications of the actual culprit (22% saved X 44% chance of identifying the actual culprit in a second sequential lineup).

This is yet another advantage of the sequential lineup. We cannot estimate exactly how many identifications of the culprit should be credited back to the sequential lineup because we cannot estimate how often police go on to find the actual culprit following a non-identification of an innocent suspect. But this advantage should not go unnoticed and it only goes in one direction; it narrows or eliminates the gap between the simultaneous and sequential lineups in the rate of culprit identifications in the real world. Consequently, the sequential advantage should be even greater in actual investigations than the meta-analysis results suggest.

Are Lost "Culprit Identifications" True Identifications?

We return now to what is perhaps the most salient issue for law enforcement and critics of the sequential lineup: the estimated eight percent drop in culprit identifications with the sequential lineup. From a policy perspective, interpretation might matter.

Why do simultaneous lineups produce higher rates of culprit identifications? We can logically dismiss the idea that the simultaneous lineup makes memory better. As discussed earlier, the key is to understand that witnesses can and do make lineup picks without true recognition. The removal-without-replacement effect (Wells, 1993) demonstrates this very well:

when a culprit is removed from a simultaneous lineup and not replaced, a large share of witnesses simply shifts their identification to another lineup member. Hence, a large share of culprit identifications are not actual recognitions of the culprit but rather simple lineup picks. The difference is important. Had witnesses used true recognition, they would have recognized the **absence** of the culprit when he was removed. If a witness picks the culprit when he is present and picks someone else when he is not present, was the pick ever really a true identification?

This brings us to an interpretation regarding higher simultaneous culprit identifications that should not be summarily dismissed. The meta-analysis provides reliable evidence of a higher choosing rate from culprit-absent simultaneous (54%) versus sequential (32%) lineups, evidence that indicates that guessing is more common with simultaneous lineups. Penrod (2003) has presented a compelling argument that guessing is a significant component of eyewitness decisions, particularly with the simultaneous lineup. Furthermore, Penrod argues that it is likely that with a simultaneous lineup, guesses "load up" on the culprit when the culprit is present because the culprit often stands out (almost no six-member lineups have a true functional or effective size of six).

So, we are left wondering about a witness who would be able to "identify" the culprit from a simultaneous lineup but could not do so with the identical lineup presented sequentially. Is this a loss of an *identification* or loss of a guess? The underlying meaning of eyewitness *identification* within the criminal justice system is that the witness "recognizes" a person based on a reliable memory. Identification means *recognition*. To the extent that the higher rate of culprit identifications with the simultaneous lineup is due to lucky guesses, it is not appropriate to call these “identifications” at all. It would be more appropriate to call them choices or picks or selections.

Our general view of why the sequential procedure disproportionately reduces mistaken identifications compared to accurate identifications is that a significant share of mistaken identifications is due to witnesses’ superficial judgments rather than true recognition. A primary basis for this superficial decision is that someone in the lineup looks more like the culprit than the others (relative judgment). Our view is that the sequential procedure helps to suppress this type of superficial lineup decision. Furthermore, it is logical to suggest that the "lost" eight percent are superficial choices—guesses.

Lost “Identifications” Might be Lower in Actual Practice

The recommended procedure (Lindsay & Wells, 1985) is that the lineup presentation continues through the entire display. The majority of researchers in the full design dataset employed this practice and in these tests the simultaneous (culprit-present lineup) advantage is reduced to five percentage points. Hence, those studies that used the stopping rule contributed substantially to the sequential lineup’s lower rate of identifying the culprit. Accordingly, there is reason to believe that the better estimate of a sequential/simultaneous difference in rates of culprit identification is five percent, not eight percent, in jurisdictions that have adopted the sequential lineup.

In actual practice, we know of no jurisdiction that has employed a stopping rule for the sequential lineup. Instead, detectives favor showing the entire array, to avoid the appearance of a show-up if the first lineup member is selected or of a truncated lineup if the procedure is terminated at an early identification. Also, if a witness selects an early filler photo, the investigating detective is later reasonably going to ask, “What would have happened when the witness saw the suspect's photo?" If a witness revokes an earlier filler pick in favor of the suspect in a later lineup position, or conversely, if the witness first picks the suspect and then discredits this initial identification with a change to a filler, this is important and useful information about the witness's memory, the quality of the lineup fillers, and about the investigative hypothesis that the suspect is the true culprit. Whether a witness who changes from a filler identification to a suspect identification is good enough evidence to survive a suppression motion from the defense is something for the courts to decide. At the least, however, this identification would likely have investigative value. Thus, the full lineup display and a record of witness remarks are important to the investigation.

Failure to Identify the Culprit Does Not Necessarily Set the Culprit Free

We have explained how lucky guesses and relative judgment may contribute to the eight percent loss in identifications of the culprit. But, what is the legal and societal cost of an eyewitness who does not identify the culprit? It is important to keep in mind that the failure of an eyewitness to identify the culprit does not automatically mean that the guilty party goes free. Guilty people tend to have other (non-witness) evidence against them and large numbers of successful prosecutions of guilty persons occur in the absence of eyewitness identification every day. Fingerprints, possession of stolen goods, confessions, semen, hair, fibers, surveillance footage, statements of co-defendants, and other types of evidence are commonly present and used against guilty people when witnesses cannot identify them. It is erroneous to think that a guilty person will go free just because a given witness failed to identify him.

Furthermore, multiple-witness cases are common (Clark & Wells, 2008). That means that any given witness might fail to identify the culprit, but the other witnesses will also have a chance. Suppose each of three witnesses separately views a culprit-present lineup. If the sequential hit rate were 44%, then the chances would be 81% that one or more witnesses would identify the culprit. Of course it is also true that a mistaken identification of an innocent person does not *necessarily* result in prosecution (and conviction). This is true for a number of reasons, including lack of corroborating evidence, a confession from another individual, or a DNA test that later excludes the innocent suspect. But even when an innocent person is not "successfully" prosecuted, his or her life can be substantially damaged by jail time, the cruel hammer of an indictment, sometimes lingering doubts of family, friends, neighbors, and co-workers, the relationship contamination of innuendo within one's social circle, and the general expenditure of time, money, and disruption of life that occurs trying to defend oneself against a mistaken identification.

Can Failures to Identify the Culprit be Equal to Identifications of an Innocent Suspect?

We acknowledge that the relative importance of the two errors, namely a mistaken identification compared to a failure to identify the culprit, is a value judgment and is not something for scientists to decide. We are intrigued, however, by a statement of Malpass, Tredoux, and McQuiston-Surrett (2009b): "We reject the idea that false identifications are necessarily more valuable for society to reduce than are correct identifications to achieve" (p.25). As a value statement, we respect the right of Malpass et al. to hold that view. But, we wonder if there is a logical rebuttal to this claim that is not based on a value judgment but instead on something more akin to math. Specifically, we note that a mistaken identification is always two errors—an identification of an innocent person and a failure to identify the culprit. Hence:

Equation 1: *Mistaken identification = inculpate the innocent + culprit escapes detection*

However, a failure to identify the culprit, in contrast, is only one error – a failure to identify the culprit. Hence:

<u>Equation 2</u>: *Failure to identify the culprit = culprit escapes detection.*

So, the culprit-escapes-detection cost applies to both errors (and hence, cancels on the two sides), but only the mistaken identification has the added feature of inculpating the innocent. Accordingly, unless one places no negative value on inculpating the innocent, the two errors cannot be equally bad and the mistaken identification has to have more negative value.

The difference between these two types of errors might even be greater than Equations 1 and 2 imply because a mistaken identification often leads the investigation away from the actual culprit, whereas a mere failure to identify the culprit does not necessarily set him free – it might simply lead to a search for other evidence. This analysis still leaves the Blackstonian question of how many of one error are equal to some number of the other error; but our analysis logically suggests that the two errors cannot be equal and that the mistaken identification must be the greater of the two.

Final Remarks

We hesitate to end our discussion with an analysis of the relative damage of the two errors (previous section) for fear that readers will think that our conclusion about sequential-superiority rests on an assumption that mistaken identifications are more damaging than failures to identify the culprit. In fact we never used that assumption and it is not needed. Instead, sequential-superiority rests solely on the observation that the sequential procedure yields a higher diagnosticity ratio and, hence, a higher posterior probability of guilt when the suspect is identified. Furthermore, the higher posterior probability of guilt holds across all possible base rates (between 0 and 100%) for the culprit being in the lineup. The data indicate that the sequential is a more rigorous test, a higher standard, and when the witness identifies the suspect, the results can be better trusted.

Numerous jurisdictions have weighed the alternatives in recent years and chosen the sequential lineup. These include the states of New Jersey, North Carolina, Ohio, and Wisconsin as well as many major cities and their counties, such as Boston, Tampa, Minneapolis/St. Paul, Dallas, Denver, and large numbers of smaller jurisdictions such as Virginia Beach (VA), Richardson (TX), Clinton (IA), and many others too numerous to name. The common element in these jurisdictions seems to be the interest in raising the probative value of identification evidence. Law enforcement wants to know that when an identification of the suspect is obtained, it is trustworthy. This is not surprising given that 75% of DNA exoneration cases trace back to mistaken identification. There is an increased understanding that eyewitness identification evidence has less probative value than previously thought. Those who have examined the DNA exoneration cases perhaps better understand now that a mistaken identification, once made, is almost impossible to distinguish from an accurate identification. Hence, the key is to keep mistaken identifications from occurring in the first place. These jurisdictions want to be able to tell the public and the pool of potential jurors that their procedures are designed to maximize the chances that an eyewitness identification is accurate.

Other jurisdictions might be motivated more by a desire to make sure that a guilty person does not escape detection (despite evidence that about 50% of guilty suspects in lineups are not identified even with the simultaneous lineup). Jurisdictions who favor this approach are likely to be less focused on diagnosticity and more focused on the eight percent (or five percent when the stopping rule is not used) reduction in identifications of the culprit. Perhaps those jurisdictions also believe that they can somehow "catch" these mistaken identifications some other way before they result in wrongful conviction. For those jurisdictions, the simultaneous lineup may be the preferred choice.

It is not for science to decide which of these policy interests is preferred. But, the data are increasingly clear and stable about the nature of that choice. Undoubtedly there are improvements to be made in lineup procedure. The sequential lineup is not intended to be the final word, as eyewitness errors still occur with this protocol. This meta-analysis has focused on the system variable of sequential lineup presentation, with an intention to provide information and guidance for sound practice—aspects of lineup construction and delivery that can be readily used by law enforcement. In that tradition, we wish to point out that research firmly establishes that mistaken identifications only occur when a lineup does not include the culprit. It is very difficult for a witness to recognize the absence of the culprit, even when a cautionary instruction is provided. The sequential lineup is a procedure that helps to reduce the risk of a dangerous false identification when the culprit is not in the lineup. However, additional strategies that can reduce the chance that an eyewitness will even encounter a culprit-absent lineup will move research and practice in a positive direction; that is, we hope to see future efforts that focus on the question of how an innocent suspect ends up in a lineup in the first place and how to minimize such occurrences (Wells, 2006).

Footnotes

1. Excluded tests: Lindsay, Lea, and Fulford (1991), Condition 3 of Exp 1 and Exp 2, involving repeated lineups; Steblay, et al., (2010), data beyond one lineup lap; MacLin & Phelen (2007) beyond one lineup lap; Vanderwal (1996), Jacob (1994), and Laldin (1997), which involved multiple culprits in one lineup; Lindsay and Bellinger (1999) in which the witness controlled the sequential lineup and 40%+ violated the no side-by-side comparison rule; Wright, Boyd, and Tredoux's (2001) within-subject comparisons for same lineup; Searcy, Bartlett, and Memon (2000) memory for a non-event in a third lineup/the data could not be separated out for effective comparison.

2. One exception is Morgan et al., (2004). The comparison of sequential to simultaneous lineups occurred across a series of studies, thereby potentially confounding a number of factors. For the purpose of the meta-analysis and to minimize confounds, we included only one Morgan comparison: between the low-stress sequential photo condition of Study 4 and the low-stress simultaneous photo condition of Study 2, both of which used single-suspect lineups. A confound still exists, in that Study 4 included both "uncued" and "cued" stimuli (clothing worn by the target), a factor manipulated in that study, while Study 2 does not include cued targets. Also, it is not clear whether the targets for both studies were the same. We include Morgan et al. only in the overall 72-test analyses. The exclusion of this test from the overall analysis (Table 1) does not change the results.

3. Filler and no-choice figures for culprit-present lineups are based on a smaller set of tests in which the frequencies were available; therefore, the tabled percentages do not add to 100%. Also, a word of caution: the absolute frequencies are the product of laboratory scenarios and not meant to convey rates that translate directly to field practice. The differences obtained between

the conditions (effect size) and the stability of that difference (*Zma*) are the more relevant basis for understanding eyewitness performance differences.

4. The finding of McQuiston-Surrett et al.—a reduction in false positives that "almost perfectly balance" (p. 141) reduction in positive identifications in the non-Lindsay lab—is not apparent in this current and larger set of data.

5. Lindsay's work with adult witnesses is distributed across published (13 tests) and unpublished (9 tests) categories, and, echoing other labs, his unpublished work reveals a diminished comparative advantage of one format over the other.

6. Two tests (Clark & Davey, 2005) use a culprit-absent lineup of size five (a culprit-removed rather than a culprit-replaced design); three tests (Lindsay et al., 1991a, 1991b) use lineups of size eight. If denominator corrections for lineup sizes (magnitude of .02) are introduced into diagnosticity calculations, the resulting numbers do not change.

7. Tests for heterogeneity of effect size indicate significant heterogeneity for all three primary dependent measures: correct identifications, correct rejections of the culprit-absent lineup, and false identifications of a designated innocent suspect, all ps < .05.

Studies in Meta-Analysis

* full diagnostic design dataset

Beaudry, J., Mansour, J., Bertrand, M., & Lindsay, R. C. L. (2006, March). *The element of surprise: The impact of participants' knowledge of a subsequent lineup.* Paper presented at the annual meeting of the American Psychology-Law Society, St. Petersburg, Florida.

Blank, H. & Krahe, J. (2000). *The influence of an outstanding similarity between two persons in a lineup on target identification in sequential and simultaneous lineups.* Unpublished manuscript. University of Leipzig, Germany.

*Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 14,* 118-128. doi:10.1037/1076-898X.14.2.118

*Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29,* 151-172. doi:10.1007/s10979-005-2418-7

*Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281-290. doi:10.1037/0021-9010.73.2.281

Dormer, G. (1983). *Effect of absolute judgmental processes on eyewitness identification accuracy.* Unpublished undergraduate thesis, Queen's University.

Douglass, A. Bradfield, & McQuiston-Surrett, D. E. (2006). Post-identification feedback: Exploring the effects of sequential photospreads and eyewitnesses' awareness of the identification task. *Applied Cognitive Psychology, 20,* 991-1007. doi:10.1002/acp.1253

Dysart, J. E. (1999). *The effects of delay on eyewitness identification accuracy.* Unpublished Masters thesis. Queens University.

Dysart, J. E., & Lindsay, R. C. L. (2001). A pre-identification questioning effect: Serendipitously increasing correct rejections. *Law and Human Behavior, 25,* 155-165. doi:10.1023/A:1005693330921

Ferch, R., & Ebbesen, E. (2003). *A signal detection application to simultaneous and sequential lineups*. Unpublished manuscript, University of California, San Diego.

Gaitens, K. E., Zimmerman, L. A., McQuiston, D. E., & Malpass, R. S. (2002). *Structural bias in eyewitness lineups: When does the suspect "stand out"?* Poster presented at the American Psychology-Law Society biennial conference, Austin, TX.

*Greathouse, S., & Kovera, M. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior, 33,* 70-82. doi:10.1007/s10979-008-9136-x

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15*, 140-152. doi:10.1037/a0015082.

Hannaford, K. M. (1985). *Relative versus absolute judgments in lineup identifications.* Unpublished undergraduate thesis, Queen's University.

*Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology, 15,* 659-671. *[This article was in the previous meta-analysis as: Kneller, W., & Memon, A. (2000). Decision processes of accurate and inaccurate eyewitnesses. Paper presented at the American Psychology-Law biennial conference, New Orleans, LA].*

*Levi, A. M. (2006). An analysis of multiple choices in MSL lineups and a comparison with simultaneous and sequential ones. *Psychology, Crime, and Law, 12,* 273-285. doi:10.1080/10683160500238782

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84,* 315-321. doi:10.1037/0021-9010.84.3.315

*Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76,* 741-745. doi:10.1037/0021-9010.76.5.741

*Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76,* 796-802. doi:10.1037/0021-9010.76.6.796

Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior, 18,* 527-541. doi:10.1007/BF01499172

*Lindsay, R. C. L., Pozzulo, J., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21,* 391-402. doi:10.1023/A:1024807202926

*Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556-564. doi:10.1037/0021-9010.70.3.556

*MacLin, O. H., & Phelen, C. M. (2007). PC_Eyewitness: Evaluating the New Jersey method. *Behavior Research Methods, 39*, 242-247.

*MacLin, O. H., Zimmerman, L. A., & Malpass, R. S. (2005). PC_Eyewitness and the sequential superiority effect: Computer-based lineup administration. *Law and Human Behavior, 29,* 303-321. doi:10.1007/s10979-005-3319-5

Martins, S. (1996). *The effects of changes in the appearance and lineup position of targets on eyewitness identification.* Unpublished undergraduate thesis, Queen's University. Data were also provided by Ferguson (1981), unpublished undergraduate thesis, Queen's University.

*Melara, R. D., Dewitt-Rickards, T. L., & O'Brien, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology, 74,* 706- 713. doi:10.1037/0021-9010.74.5.706

Memon, A., & Bartlett, J. (2002). The effects of verbalization on face recognition in young and older adults. *Applied Cognitive Psychology, 16,* 635-650. doi:10.1002/acp.820

*Memon, A., & Gabbert, F. (2003a). Improving the identification accuracy of senior witnesses: Do prelineup questions and sequential testing help? *Journal of Applied Psychology, 88,* 341-347. doi:10.1037/0021-9010.88.2.341.

Memon, A., & Gabbert, F. (2003b). Unravelling the effects of sequential presentation in culprit present lineups. *Applied Cognitive Psychology, 17,* 703-714. doi:10.1002/acp.909

Morgan, III. C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P., Baranoski, M., Southwick, S.M. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry, 27,* 265-279. doi:10.1016/j.ijlp.2004.03.004

Newman, K. (1998). *The effects of a weak memory trace on sequential and simultaneous lineup identifications.* Unpublished undergraduate thesis, Queen's University.

Ogle, C. M., & Reisberg, D. (2006). *A comparison of elimination, sequential, and simultaneous lineup procedures.* Paper presentation at Association for Psychological Science convention, New York, NY.

*Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. *Law and Human Behavior, 17,* 11-26. doi:10.1007/BF01044534

Parker, J. F., Tredoux, C., & Nunez, D. (2000). *Lineup measures, lineup procedure, and optimality of encoding.* Paper presented at the American Psychology-Law Society biennial conference, New Orleans, LA.

Phillips, M. R., McAuliff, B. D., Bull Kovera, M., & Cutler, B. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84,* 940-951. doi:10.1037/0021-9010.84.6.940

*Pozzulo, J. D., Corey, S. C., Kiraly, M., Girardi, A., Lawandi, A., & Aston, C. (2008). Can a lineup procedure designed for child witnesses work for adults: Comparing simultaneous, sequential, and elimination lineup procedures. *Journal of Applied Social Psychology, 38,* 2195-2209. doi:10.1111/j.1559-1816.2008.00387.x

*Pozzulo, J. D., & Marciniak, S. (2006). Comparing identification procedures when the perpetrator has changed appearance. *Psychology, Crime, and Law, 12,* 429-438. doi:10.1080/10683160500050690

Rombough, V. J. (1994). *The effects of lineup presentation and practice on lineup identification accuracy in child eyewitnesses.* Unpublished undergraduate thesis, Queen's University.

*Rose, R. A., Bull, R., & Vrij, A. (2005). Non-biased lineup instructions do matter – a problem for older witnesses. *Psychology, Crime, and Law, 11,* 147-159. doi:10.1080/10683160512331316307

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older witnesses. *Legal and Criminological Psychology, 5,* 219-235. doi:10.1348/135532500168100

Shapiro, L. R., & Hiatt, T. A. (2002). *A comparison of photographic line-up procedures.* Paper presented at the American Psychology-Law Society biennial conference, Austin, TX.

Smith, S. M., Lindsay, R. C. L., Pryke, S., & Dysart, J. E. (2001). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross race situations? *Psychology, Public Policy, and Law, 7,* 153-169. (unpublished sequential lineup data secured from S. Smith). doi:10.1037/1076-8971.7.1.153

Smyth, L. (1994). *Sequential presentation and practice: The jury is still out.* Unpublished undergraduate thesis, Queen's University.

*Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78,* 22-33. doi:10.1037/0021-9010.78.1.22

*Steblay, N. K., Dietrich, H. L., Ryan, S. L., Raczynski, J. L., & James, K. A. (in press). Sequential lineup laps and eyewitness accuracy. *Law and Human Behavior.*

Varrette, T. K. (1994). *The effects of similarity-to-suspect versus match-to-description strategies for selecting lineup foils when perpetrator appearance has been altered*. Unpublished undergraduate thesis, Queen's University.

*Wells, E. C., & Pozzulo, J. D. (2006). Accuracy of eyewitnesses with a two-culprit crime: Testing a new identification procedure. *Psychology, Crime, & Law, 12,* 417-427. doi:10.1080/10683160500050666

*Wilcock, R. A., Bull, R., & Vrij, A. (2005). Aiding the performance of older eyewitnesses: Enhanced non-biased line-up instructions and line-up presentation. *Psychiatry, Psychology and Law, 12,* 129-140. doi:10.1375/pplt.2005.12.1.129

Yarmey, A. D., & Morris, S. (1998). The effects of discussion on eyewitness memory. *Journal of Applied Social Psychology, 28,* 1637-1648. doi:10.1111/j.1559-1816.1998.tb01693.x

References

Blank, H., & Krahe, J. (2000). *The influence of an outstanding similarity between two persons in a lineup on target identification in sequential and simultaneous lineups.* Unpublished manuscript. University of Leipzig, Germany.

Charman, S., & Wells, G. L. (2006). Applied lineup theory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology, Volume 2* (pp. 219-254). New Jersey: Lawrence Erlbaum.

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29,* 151-172. doi:10.1007/s10979-005-2418-7

Clark, S. E., Erickson, M. A., & Breneman, J. S. (in press). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*.

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 3,* 187-218. doi:10.1007/s10979-006-9082-4.

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 32*, 406-422. doi:10.1007/s10979-007-9115-7.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S.Ct. 2786 (1993).

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28,* 687-706. doi:10.1007/s10979-004-0565-x

Gaertner , S., & Harrington, J. (2009). Successful eyewitness identification reform: Ramsey County's blind sequential lineup protocol. *The Police Chief, 76*, 130-141.

Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied, 15*, 140-152. doi:10.1037/a0015082.

Jacob, P. (1994). *The feasibility of using multiple perpetrator sequential lineups.* Unpublished undergraduate thesis, Queen's University.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697-720. doi:10.1037/0003-066X.58.9.697

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291. doi:10.2307/1914185

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

Laldin, S. (1997). *Contextual effects on lineup identification of multiple perpetrators*. Unpublished undergraduate thesis, Queen's University.

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84,* 315-321. doi:10.1037/0021-9010.84.3.315

Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76,* 741-745. doi:10.1037/0021-9010.76.5.741

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76,* 796-802. doi:10.1037/0021-9010.76.6.796

Lindsay, R. C. L., Mansour, J. K., Beaudry, J. L., Leach, A. M., & Bertrand, M. I. (2009a). Sequential lineup presentation: Patterns and policy. *Legal and Criminological Psychology, 14,* 13-24. doi:10.1348/135532508X382708

Lindsay, R. C. L., Mansour, J. K., Beaudry, J. L., Leach, A. M., & Bertrand, M. I. (2009b). Beyond sequential presentation: Misconceptions and misrepresentations of sequential lineups. *Legal and Criminological Psychology, 14,* 31-34. doi:10.1348/135532508X382104

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556-564. doi:10.1037/0021-9010.70.3.556

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distractors for lineups. *Law and Human Behavior, 15,* 43-57. doi:10.1007/BF01044829

MacLin, O. H., & Phelen, C. M. (2007). PC_Eyewitness: Evaluating the New Jersey method. *Behavior Research Methods, 39*, 242-247.

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. E. (2009a). Public policy and sequential lineups. *Legal and Criminological Psychology, 14,* 1-12. doi:10.1348/135532508X384102

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. E. (2009b). Response to Lindsay, Mansour, Beaudry, Leach & Bertrand's 'Sequential lineup presentation: Patterns and policy'. *Legal and Criminological Psychology, 14,* 25-30. doi:10.1348/135532508X384094

McQuiston-Surrett, D. E., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of method, data, and theory. *Psychology, Public Policy and Law, 12,* 137-169. doi:10.1037/1076-8971.12.2.137

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15,* 603-616. doi:10.1001/acp.728

Memon, A., & Gabbert, F. (2003). Improving the identification accuracy of senior witnesses: Do prelineup questions and sequential testing help? *Journal of Applied Psychology, 88,* 341-347. doi:10.1037/0021-9010.88.2.341

Penrod, S. D. (2003). Eyewitness identification evidence: How well are witnesses and police performing? *Criminal Justice Magazine, 54*, 36-47.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older witnesses. *Legal and Criminological Psychology, 5,* 219-235. doi:10.1348/135532500168100

Steblay, N. K. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21,* 283-297. doi:10.1023/A:1024890732059

Steblay, N. K., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25,* 459-473. doi:10.1023/A:1012888715007

Steblay, N. K., Dietrich, H. L., Ryan, S. L., Raczynski, J. L., & James, K. A. (in press). Sequential lineup laps and eyewitness accuracy. *Law and Human Behavior.* doi:10.1007/s10979-010-9236-2

Technical Working Group for Eyewitness Accuracy. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: U.S. Department of Justice.

Vanderwal, A. (1996). *The effects of the sequential presentation of lineups with multiple perpetrators on eyewitness identification.* Unpublished undergraduate thesis, Queen's University.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 12,* 1546-1557. doi:10.1037/0022-3514.36.12.1546

Wells, G. L. (1984). The psychology of lineup identification. *Journal of Applied Social Psychology, 14,* 89-103. doi:10.1037/0022-3514.36.12.1546

Wells, G. L. (2006). Eyewitness identification: Systemic reforms. *Wisconsin Law Review, 2,* 615-643.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776-784. doi:10.1037/0033-2909.88.3.776

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist, 55,* 581-598. doi:10.1037//0003-066X.55.6.581.

Wells, G. L., & Penrod, S. D. (in press). Eyewitness identification research: Strengths and weaknesses of alternative methods. In B. Rosenfeld, & S. D. Penrod (Eds.), *Research methods in forensic psychology.* New York: Wiley.

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99,* 320-329. doi:10.1037/0033-2909.99.3.320

Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). A field study of own-race bias in South Africa and England. *Psychology, Public Policy, and Law, 7,* 119-133. doi:10.1037/0033-2909.99.3.320

Table 1

Lineup Performance: Sequential vs. Simultaneous Lineup Formats

*2010 Data (2001 Meta-analysis Data in Parentheses) * p < .05*

Eyewitness decisions	k	Sequential % 2010 (2001)	Simultaneous % 2010 (2001)	r	*(Zma)
Culprit-present lineup					
Culprit ID	58	.38 (.35)	.52 (.50)	-.14	*
Filler	48	.24 (.19)	.24 (.24)		
No choice	48	.41 (.46)	.27 (.26)		*
Culprit-absent lineup					
Correct rejection	64	.64 (.72)	.43 (.49)	.22	*
Filler	64	.36 (.28)	.57 (.51)		*
Identification of designated innocent suspect	27	.13 (.09)	.25 (.27)	.14	*

Table 2

Moderator Variable Analyses: *Robustness of the Sequential-superiority Effect*

	Culprit-present Correct IDs *r* (k)	Culprit-Absent Correct rejections *r* (k)	Innocent suspect *r* (k)
Witness age			
Adults	-.13 (48) * [a]	.24 (56) * [a]	.15 (25) * [a]
Older adults	-.25 (5) * [b]	.27 (4) *	---
Children under 12	-.19 (4) *	.00 (3) *ns* [b]	.03 (2) [b]
Lindsay lab (adult witnesses only)			
Lindsay k =22	-.13 (15) *	.28 (20) *	.26 (13) * [a]
Others k = 39	-.13 (33) *	.22 (36) *	.03 (12) * [b]
Publication status (adult witnesses only)			
Published k = 45	-.10 (34) * [a]	.25 (42) *	.15 (18) *
Unpublished k = 16	-.19 (14) * [b]	.19 (14) *	.14 (7) *
Full design dataset			
Full design dataset k = 27	-.08 (27) * [a]	.23 (27) *	.13 (11) *
All remaining tests k = 45	-.19 (31) * [b]	.21 (37) *	.15 (16) *

* difference between simultaneous and sequential lineups is statistically significant, $p < .05$

ab: superscripts that differ between groups, within condition of culprit-present or culprit-absent, signify a statistically significant difference at $p < .05$ between the effect sizes.

Not included in the listing are “mixed conditions” (in which participants within the same study experienced different levels of the variable), cases in which the variable was unreported, or conditions in which the number of tests was very small.

Table 3

Lineup Performance: Sequential vs. Simultaneous Lineup Formats

Full Diagnostic Design Subset of 2x2 Designs, Published, Adult Witnesses

Eyewitness decision	k	Sequential %	Simultaneous %	r	*(Zma)
Culprit-present lineup					
Culprit ID	27	.44	.52	-.08	*
Filler	24	.19	.25	-.07	*
No choice	24	.39	.24		*
Culprit-absent lineup					
Correct rejection	27	.68	.46	.23	*
Filler	27	.32	.54		*
Diagnosticity ratio		7.72	5.78		

Identification of designated innocent suspect					
	11	.15	.28	.13	*
Diagnosticity ratio		2.94	1.86		

Table 4

Moderator Analysis: System Variables in Sequential and Simultaneous Lineups (Full design dataset)

					Diagnosticity	
	k	CPr	CAr	DSr (k)	SEQ	SIM
Lineup construction method						
Match to description	15	-. 10 *	.24 *	.16 (8) *	10.00	6.64
Match to culprit	9	-.06 *	.23 *	.06 (3) *ns*	5.21	3.82
Back-loading						
Back-loaded	22	-.09 *	.26 *	.13 (9) *	8.18	5.47
Not back-loaded	4	.01 *ns*	.13 *	.12 (2)	8.53	6.00
Description of culprit before lineup						
Yes	14	-.05 *	.21 *	.14 (8) *	8.68	5.93
Unreported/no	13	-.12 *	.26 *	.11 (3) *	7.65	5.51
Stopping rule for sequential lineup						
Stop at ID	6	-.17 * [a]	.28 *	.08 (1)	8.43	6.34
Continue to end of lineup	20	-.05 * [b]	.19 *	.14 (10) *	8.34	5.92
Treatment of multiple IDs (for studies that continue the sequential lineup)						
None made multiple IDs	4	-.12 *	.23 *	---	6.54	5.06
Counts as false alarm	4	-.12 *	.17 *	. 23 (3) * [a]	6.21	5.29
1st response rule	9	- .03*ns*	.18 *	.03 (5) * [b]	8.42	5.86

Moderator analyses: laboratory method

Position of suspect				
Full rotation	10	-.13 *	.22 *	.12 (3) *
Partial rotation	5	-.02 *ns*	.23 *	.06 (4)
Stable in position (3+)	5	-.12 *	.25 *	.15 (2) *
Photo order counterbalanced/randomized				
Counterbalanced	13	-.10 *	.21 *	.05 (6) *
No counterbalance	5	-.12 *	.23 *	.19 (2) *
Control of experimenter expectancy effects				
Blinded	21	-.09 *	.26 *	.07 (6) *
Not blind in at least 1/2	5	-.04 *	.23 *	.20 (5) *
Presence of designated innocent suspect in culprit-absent lineup				
Designated suspect	11	-.05 *	.20 *	.13 *
No designated suspect	15	-.10 *	.26 *	---
Lineup fairness check				
Ranked visual similarity	7	-.04 ns	.17	.10 (3) ns
Mock witness procedure	2	-.15 *	.26 *	---
Functional/effective size	6	-.06 *	.13 *	.09 (5) *

k = number of tests

CPr = effect size r for Culprit-Present condition

CAr = effect size r for Culprit-Absent condition

DSr (k) = effect size r for Designated Suspect; (k)= number of tests

Diagnosticity for SEQ (sequential) and SIM (simultaneous) lineups

ns = a not-significant finding, but also one in which obtained effect sizes are on both sides of zero; i.e., a “zero” effect hides findings in which each of sequential and simultaneous lineups are at times favored.

Figure Captions

Figure 1. Culprit present lineups: Stem and leaf of effect sizes (r), k = 58 tests (Full design data in **bold**).

Figure 2. Culprit absent lineups: Stem and leaf display of effect sizes (r), k = 64 tests (Full design

data in **bold**)

Figure 3. Posterior (after identification) probabilities that the suspect is the culprit across all possible prior (base rate) probabilities that the suspect is the culprit as a function of simultaneous and sequential procedures for the full design dataset. Identity line represents the posterior probability if an identification had no diagnostic value. Difference refers to the amount of difference favoring the sequential procedure.

Figure 4. Posterior (after identification) probabilities that the suspect is the culprit across all possible prior (base rate) probabilities that the suspect is the culprit as a function of simultaneous and sequential procedures for the designated innocent suspect dataset. Identity line represents the posterior probability if an identification had no diagnostic value. Difference refers to the amount of difference favoring the sequential procedure.

Figure 1.

Stem	Leaf
.8	
.7	
.6	
.5	
.4	
.3	
.2	**1**
.1	**0**,**1**
.0	**0**,0,0,**5**,5,**6**,**7**
-.0	**3**,4,4,**5**,**6**,6,6,**8**,**8**,**8**,8,8,**9**,9
-.1	0,**0**,**0**,**0**,**1**,**1**,1,3,**5**,**6**,8,**9**
-.2	0,0,1,**2**,2,4,6,**6**,**8**,9,9,9
-.3	**0**,0,0,1,**2**,2,4,5
-.4	3,7
-.5	
-.6	
-.7	
-.8	

Figure 2.

.8	
.7	**8**
.6	1
.5	4,7
.4	**1**,1,2,3,3,**5**,**6**,**6**,6,7
.3	**0**,0,**2**,3,3,**4**,**6**,6,6
.2	**0**,0,1,**2**,2,**3**,**3**,**3**,**9**,**9**,9
.1	**0**,0,0,1,2,4,6,7,**9**
.0	**2**,**2**,**2**,2,2,4,4,4,**8**,**8**,8,**9**,**9**,**9**,9,9
-.0	**4**,5,5,7
-.1	4
-.2	
-.3	
-.4	
-.5	
-.6	
-.7	
-.8	

Figure 3.

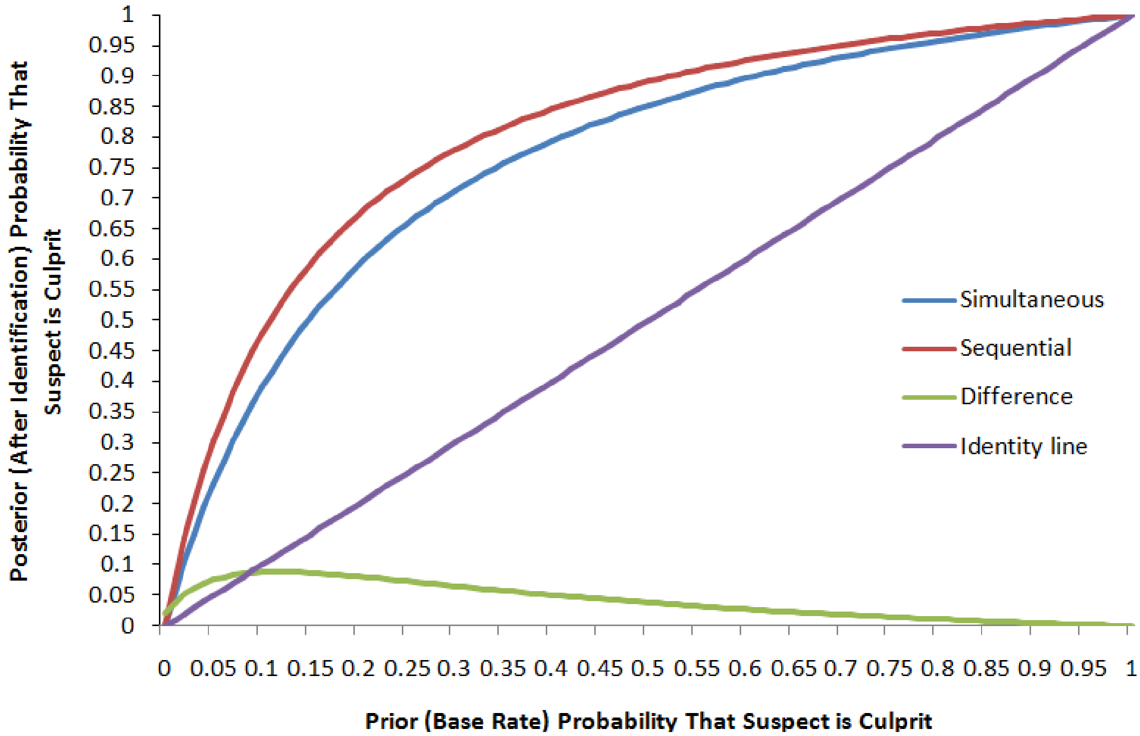
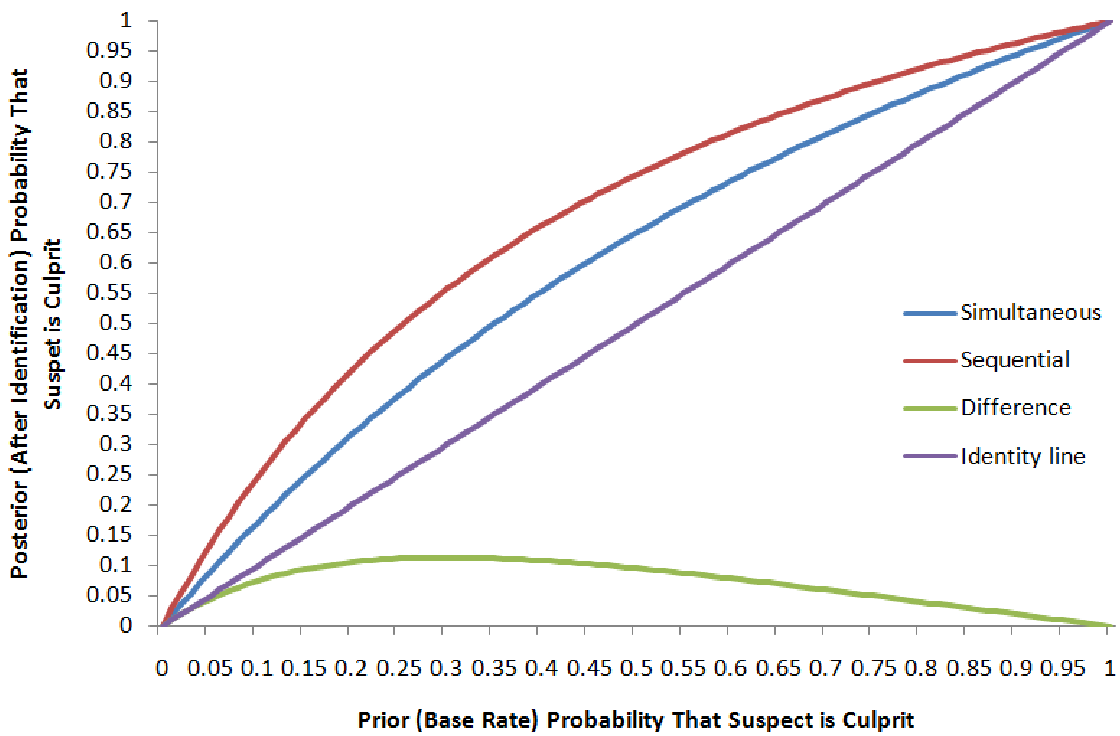

Posterior (After Identification) Probability That Suspect is Culprit
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
Prior (Base Rate) Probability That Suspect is Culprit
Simultaneous
Sequential
Difference
Identity line

Figure 4.


Posterior (After Identification) Probability That Suspet is Culprit
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
Prior (Base Rate) Probability That Suspect is Culprit
Simultaneous
Sequential
Difference
Identity line

Appendix A

Studies in the Meta-Analytic Calculations (72 tests)

		Culprit Present r	Culprit Absent r	Inno	Diag	Adult	LL	Pub
Beaudry Mansour Bertrand Lindsay	2006	-.06	.04			A	L	
Blank & Krahe	2000	-.43				A		
Carlson Gronlund & Clark	2008	-.16	.09	I	D	A		P
(2)		-.11	.20	I	D	A		P
Clark & Davey	2005	.21	.02	I	D	A		P
(2)		.06	.02	I	D	A		P
Cutler & Penrod	1988	.05	.23		D	A		P
(2)		-.06	.22		D	A		P
Dormer	1983	-.08	.16	I		A	L	
Douglass & McQuiston-Surrett	2006		.21			A		P
(2)		-.31	.02			A		P
Dysart	1999	-.04	.10			A	L	
Dysart & Lindsay	2001		.41			A	L	P
Ferch & Ebbesen	2003	-.29	.22			A		
(2)		-.47	.30	I		A		
Gaitens et al.	2002		.17	I		A		
Greathouse & Kovera	2009	.00	.08	I	D	A		P
Gronlund, et al	2009	-.10	.14	I		A		P
Hannaford	1985	-.21	.02	I		A	L	

Kneller Memon & Stevenage	2001	-.11	.41		D	A		P
Levi	2006	-.28	.19		D	A		P
Lindsay & Bellinger	1999		.43			A	L	P
Lindsay Lea & Fulford	1991	-.10	.34	I	D	A	L	P
(2, from Exp 3)			.09			A	L	P
Lindsay et al.	1991	.11	.30	I	D	A	L	P
(2)		-.10	.36	I	D	A	L	P
(3)			.46	I		A	L	P
(4)			.36	I		A	L	P
(5)			.57	I		A	L	P
Lindsay Martin & Webber	1994		.33	I		A	L	P
Lindsay et al.	1997	.07	.10		D	A	L	P
(2)		-.11	-.14			T	L	P
(3)		-.06	-.07			C	L	P
(4)		-.32				A	L	P
(5)		-.26				C	L	P
Lindsay & Wells	1985	-.08	.23	I	D	A	L	P
MacLin & Phelen	2007	-.26	.45	I	D	A		P
MacLin Zimmerman & Malpass	2005	-.08	.23		D	A		P
(2)		-.22	.29		D	A		P
Martins/Ferguson	1996	-.20				A	L	
Melara Dewitt-Rickards & O’Brien	1989	-.15	.78		D	A		P
Memon & Bartlett	2002	-.18				A		P

(2)		-.30				O		P
Memon & Gabbert	2003a	-.24				A		P
(2)		-.09				O		P
Memon & Gabbert	2003b	-.32	.46		D	A		P
(2)		-.29	.54			O		P
Morgan et al.	2004	.00	.61			A		P
Newman	1998	-.29	.11			A	L	
Ogle & Reisberg	2006	.00	.33			A		
Parker & Ryan	1993	-.09	.09	I	D	A		P
(2)		-.13	-.04	I		C		P
Parker Tredoux & Nunez	2000		.12			A		
Phillips et al.	1999		-.05	I		A		P
Pozzulo et al.	(in press)	-.08	.29		D	A		P
Pozzulo & Marciniak	2006	-.10	.09		D	A		P
Rombough	1994	-.30	.10	I		C	L	
Rose Bull & Vrij	2005	-.30	.08		D	A		P
(2)		-.22	.09			O		P
Searcy Bartlett & Memon	2000		.42			A		P
(2)			.36			O		P
Shapiro & Hiatt	2002	.05	-.05			A		
Smith Lindsay Pryke & Dysart	2001	-.20	.20	I		A	L	
Smyth	1994	-.34	.43	I		A	L	
Sporer	1993	-.05	.32		D	A		P

Steblay	2010	-.04	.04			A		P
(2)		.10	-.04		D	A		P
Varrette	1994	-.08	.47	I		A	L	
Wells & Pozzulo	2006	-.19	.02		D	A		P
Wilcock Bull & Vrij	2005	-.03	.46		D	A		P
(2)		-.35	.08			O		P
Yarmey & Morris	1998		.29	I		A		P
		58	64	27	27		27	55

Culprit-Present r: Effect size for correct identifications

Culprit-Absent r: Effect size for correct rejections

Inno: tests that include a designated innocent suspect (27)

Diag: the Full design Dataset—published tests, adult witnesses, 2 X 2 design (29).

Adult: age category of the witness-participants (Age: A = Adult; O = Older adult; C = Children under 12; T = teen)

LL: tests from the Lindsay lab (27)

Pub: Published tests (55)

Appendix B

Study Features of the Full Diagnoticity Design Dataset (27 tests)

		Lineup Const	Fair	Perp Position	ExpE	Stop	Photo Order	Back-loaded
Carlson Gronlund & Clark	2008	Desc	size	Full	Multi	End	Bal	Back
(2)		Desc	size	Part	Multi	End	Bal	Back
Clark & Davey	2005	Culp	visual	Part	Blind	End	Bal	Back
(2)		Culp	visual	Part	Blind	End	Bal	Back
Cutler & Penrod	1988	--	--	--	Multi	End	--	Back
(2)		--	--	--	Multi	End	--	No
Greathouse & Kovera	(in press)	Desc	Size	Fixed	Mixed	Stop	No	No
Kneller et al.	2001	Culp	Visual	Full	--	Stop	--	Back
Levi	2006	--	--	--	Multi	Stop	Bal	--
Lindsay Lea & Fulford	1991	Desc	Size	Fixed	Mixed	End	--	Back
Lindsay et al.	1991	Desc	Visual	--	Mixed	End	--	Back
(2)		Desc	--	--	Mixed	End	--	Back
Lindsay et al.	1997	Desc	--	--	Multi	End	--	Back
Lindsay & Wells	1985	Desc	--	Part	Multi	End	No	Back
MacLin et al	2007	Desc	Visual	Full	Blind	End	Bal	Back
MacLin et al.	2005	Desc	Mock	--	Multi	End	--	Back
(2)		Desc	Mock	--	Multi	End	--	Back
Melara et al.	1989	Culp	--	Part	Blind	--	Bal	Back
Memon & Gabbert	2003b	Desc	--	Fixed	Multi	Stop	No	Back

Parker & Ryan	1993	Culp	Size	Full	Mixed	End	Bal	No
Pozzulo et al.	2008	Culp	--	Fixed	Multi	End	No	Back
Pozzulo & Marciniak	2006	Culp	--	Fixed	Multi	End	No	Back
Rose Bull & Vrij	2005	Desc	Visual	Full	Multi	Stop	Bal	Back
Sporer	1993	Culp	Visual	Full	Multi	End	Bal	Back
Steblay	2010	Desc	Size	Full	Multi	End	Bal	Back
Wells & Pozzulo	2006	Culp	Visual	Full	Multi	End	--	Back
Wilcock Bull & Vrij	2005	Desc	--	Full	Multi	Stop	Bal	Back

Lineup Const: Lineup construction method (Desc = Match to description; Culp = match to culprit)

Fair: Lineup fairness assessment (Size = functional or effective size calculated; Mock = mock witness method; Visual = lineup members checked for visual similarity)

Perp Position: Position of the culprit in the lineup (Full = full rotation through the lineup, except position 1; Part = partial rotation through 2-4 positions; Fixed = fixed in position)

ExpE: Control of experimenter expectancies (Multi = multiple methods; Blind = full blind; Mixed = not blind in at least half the conditions)

Stop: Stopping rule (End = continue to end of lineup; Stop = stop at an identification)

Photo Order: Lineup photos order (Bal = counterbalanced or randomized; No = no counterbalance)

Back: Back-loading (Back = back-loaded lineup, or subjects not aware of number of photos; No = no back-load)