Sequential Lineup Presentation Promotes Less-Biased Criterion Setting but Does Not Improve Discriminability

Matthew A. Palmer and Neil Brewer
Flinders University

When compared with simultaneous lineup presentation, sequential presentation has been shown to reduce false identifications to a greater extent than it reduces correct identifications. However, there has been much debate about whether this difference in identification performance represents improved discriminability or more conservative responding. In this research, data from 22 experiments that compared sequential and simultaneous lineups were analyzed using a compound signal-detection model, which is specifically designed to describe decision-making performance on tasks such as eyewitness identification tests. Sequential (cf. simultaneous) presentation did not influence discriminability, but produced a conservative shift in response bias that resulted in less-biased choosing for sequential than simultaneous lineups. These results inform understanding of the effects of lineup presentation mode on eyewitness identification decisions.

Keywords: eyewitness identification, sequential lineup advantage, signal detection theory, compound signal detection

One proposed means of improving eyewitness identification accuracy is to use a sequential lineup presentation, such that the witness views one lineup member at a time (Lindsay & Wells, 1985). As compared with simultaneous lineup presentation, sequential presentation is associated with fewer correct identifications from culprit-present lineups (i.e., lineups that contain the culprit). However, this loss in correct identifications is typically outweighed by a larger reduction in false identifications from culprit-absent lineups (i.e., lineups that do not contain the culprit), resulting in an overall gain in accuracy (Clark, Howell, & Davey, 2008; Goodsell, Gronlund, & Carlson, 2010; Steblay, Dysart, Fulero, & Lindsay, 2001; Steblay, Dysart, & Wells, 2011). For example, in a recent meta-analysis, Steblay et al. (2011) estimated that, as compared with simultaneous lineups, sequential lineups produce 22% fewer false identifications and only 8% fewer correct identifications.

There has been considerable debate about how to interpret the difference in identification performance between sequential and simultaneous lineups (e.g., Ebbesen & Flowe, 2002; Gronlund, 2004; Lindsay, Mansour, Beaudry, Leach, & Bertrand, 2009; Malpass, Tredoux, & McQuiston-Surratt, 2009; McQuiston–Surratt, Malpass, & Tredoux, 2006; Meissner, Tredoux, Parker, & MacLin, 2005; Wells, 2001). One prominent point of contention is whether the difference in identification performance between sequential and simultaneous lineups represents a difference in discriminability (i.e., witnesses’ ability to distinguish the culprit from other lineup members) or response bias (i.e., the tendency of witnesses to choose from or reject a lineup), constructs rooted in signal detection theory (SDT; Green & Swets, 1966; Macmillan & Creelman, 1991). Recently, researchers have advocated the use of quantitative models for developing an understanding of the cognitive mechanisms underpinning the sequential lineup advantage (Clark, 2008; Goodsell et al., 2010; Wells, 2008). In this research, we used a compound signal detection model to reanalyze data from 22 experiments that compared sequential and simultaneous lineups in order to investigate whether the differences in identification performance between the two presentation modes could be attributed to differences in discriminability or response bias.

The Nature of the Sequential Lineup Advantage

There are two reasons why it is important to explore the nature of the sequential lineup advantage. First, from an applied perspective, it has been argued that this issue needs to be resolved before the sequential procedure can be recommended for use in forensic settings (McQuiston–Surratt et al., 2006). Indeed, some researchers have expressed reluctance to recommend a procedure that causes a conservative shift in responding as a practical alternative to simultaneous lineups (Meissner et al., 2005). Second, the question of whether the sequential advantage reflects greater discriminability or a conservative shift in response bias holds considerable interest from a theoretical perspective. For example, a better understanding of the nature of the sequential lineup advantage would help shape the development of models of eyewitness identification decisions (e.g., WITNESS; Clark, 2003; Goodsell et al., 2010).

Prior research has not elucidated the nature of the sequential lineup advantage. Published studies that have addressed this issue...
have examined decisions that are similar to, but not the same as, eyewitness identification decisions. Gronlund (2004) had participants make a series of relative or absolute judgments about the height of target persons, and Meissner et al. (2005) used a lineup recognition paradigm, in which participants viewed a series of photographs of target persons and then attempted to identify them from a series of lineups. In addition, published (Gronlund, 2004; Meissner et al., 2005) and unpublished (Ebbesen & Flowe, 2002) studies have used models based on relatively simple decisions (e.g., “yes–no” recognition), rather than on the more complex type of decision faced by an eyewitness viewing a lineup. In our research, we used a model specifically designed to describe decision-making performance on tasks such as identification tests: SDT-compound decision (SDT-CD; Duncan, 2006; Duncan, 2011; Palmer, Brewer, & Weber, 2010; see also Starr, Metz, Lusted, & Goodenough, 1975). As compared with models based on simpler decisions, SDT-CD produces estimates of discriminability and response bias that more accurately describe decision-making performance on identification tests.

Figure 1 shows a simple SDT representation of an identification decision. The horizontal axis represents the strength of evidence in favor of making a positive identification. Note that some details are omitted for the sake of clarity; for example, there may be a separate distribution for each nonculprit lineup member (e.g., Ebbesen & Flowe, 2002) and the variance of the culprit distribution would likely be greater than that of the nonculprit distribution (Ratcliff, Sheu, & Gronlund, 1992).

The discriminability account of the sequential lineup advantage holds that sequential presentation improves witnesses’ ability to distinguish culprits from nonculprits. This equates to greater separation between the culprit and nonculprit distributions and, in turn, a higher ratio of hits (positive responses to the culprit) to false alarms (positive responses to nonculprits). In contrast, the criterion shift account is represented by a conservative shift in decision criterion such that more evidence is required for a positive identification under sequential versus simultaneous presentation conditions (depicted in Figure 1 as a shift from $C_{\text{SIM}}$ to $C_{\text{SEQ}}$). Consequently, there is a decrease in both the hit rate and false alarm rate. It is important to note that these decreases are not necessarily equal in magnitude; the relative effects on hits and false alarms depend on where $C_{\text{SIM}}$ and $C_{\text{SEQ}}$ are positioned in relation to the culprit and nonculprit distributions. The point of intersection between the culprit and nonculprit distributions represents an unbiased response criterion, indicating that the responder has no tendency to favor positive responses or negative responses (shown in Figure 1 as $C_{\text{UNBIASED}}$). In experiments, where half of lineups typically contain the target, provided discriminability has a positive value (i.e., where the culprit distribution is positioned to the right of the nonculprit distribution). For example, the shift from $C_{\text{SIM}}$ to $C_{\text{SEQ}}$ (in Figure 1) represents a shift from a criterion that is too lenient to a criterion that is less biased. In this situation, the decrease in hits will be outweighed by a larger decrease in false alarms, resulting in greater overall accuracy for sequential than simultaneous lineups.

Because a criterion shift can affect hits and false alarms differently, the criterion shift account can potentially explain the findings of numerous studies that have compared sequential with simultaneous presentation. We know that witnesses have a tendency to choose from simultaneous lineups (e.g., Wells, 1993) and that sequential presentation reduces choosing and increases overall accuracy (Clark et al., 2008; Steblay et al., 2001; Steblay et al., 2011). Therefore, it could be that sequential presentation (relative to simultaneous) produces a criterion shift from a point that is too lenient to a point that is less biased (as per Figure 1). In the debate about whether the sequential advantage rests on a discriminability or response criterion mechanism, insufficient attention has been given to the placement of witnesses’ response criterion in relation to the point of unbiased responding. As a result, the sequential presentation method has perhaps been undersold. Sequential presentation might not seem an appealing option if it merely causes

![Figure 1](image-url)
witnesses to set a more conservative response criterion. But what if it causes witnesses to set a less-biased response criterion? At this point, it is important to note that less-biased responding does not always produce higher accuracy rates. Although this is the case in experiments where there are equal numbers of culprit-present and culprit-absent lineups, it does not apply in some other situations. For example, when the base rate of culprit-present lineups is very high, accuracy is maximized when responding is lenient. Thus, although the criterion shift account described above predicts that sequential presentation will increase accuracy in experimental settings, it cannot be taken for granted that this also applies to police investigations (where the base rate of culprit-present lineups is not known). We elaborate on this issue in the Discussion.

Estimating Discriminability and Response Bias for Compound Decisions

The most common approach for teasing apart differences in discriminability from differences in response bias is to calculate indices that reflect these two parameters (e.g., $d'$ and $c$ for discriminability and bias, respectively). For simple decisions (e.g., old/new recognition judgments) where only four responses are possible (hits and misses for target trials, and false alarms and correct rejections for distractor trials) $d'$ and $c$ can easily be calculated from hit and false alarm rates. However, calculating these indices is problematic for eyewitness identification decisions, because a fifth type of response is possible: foil identifications from culprit-present lineups. The difficulty arises because these responses could be classified as both false alarms (because a foil was chosen) and misses (because the target was present but not chosen). Due to a lack of formal models that accommodate foil identifications from culprit-present lineups (Duncan, 2006, Duncan, 2011), previous studies that have used a signal detection approach to compare identification performance between simultaneous and sequential lineups have relied on proxy estimates of discriminability that take into account foil identifications from target-absent lineups but not target-present lineups (Ebbesen & Flowe, 2002; Gronlund, 2004; Meissner et al., 2005).

For many analyses of identification responses, it is appropriate to exclude foil identifications from culprit-present lineups because these responses represent known errors in the sense that they involve choices of known-to-be-innocent lineup members (e.g., Brewer & Wells, 2006; Sauer, Brewer, Zweck, & Weber, 2010). However, for the purposes of describing the decision-making performance of eyewitnesses, foil identifications from culprit-present responses must be taken into account. Consider two hypothetical lineup procedures, A and B, which produce identical rates of correct identifications from culprit-present lineups (say, 40%) and false identifications of innocent suspects from culprit-absent lineups (20%). The known-to-be-innocent foil identification rate for culprit-present lineups is 60% for procedure A and 0% for procedure B. In terms of the probative value of suspect identifications, procedures A and B are identical. However, in terms of the decision-making performance of eyewitnesses, the two procedures are clearly very different.

Similarly, for culprit-absent lineups, positive identifications of a designated innocent suspect are often treated as different to positive identifications of other lineup members (e.g., Goodsell et al., 2010; Gronlund, Carlson, Dailey, & Goodsell, 2009; Lindsay & Wells, 1985). However, when the focus is on describing the decision-making performance of eyewitnesses (as is the case in this research), all positive identifications from culprit-absent lineups must be treated as incorrect positive responses. Thus, although many studies in the eyewitness literature differentiate between suspect and foil identifications for culprit-absent lineups, we do not do so here.

In this research, we used SDT-CD (Duncan, 2006, Duncan, 2011; Palmer et al., 2010; see also Starr et al., 1975), a model that considers foil identifications from culprit-present and culprit-absent lineups and, hence, produces estimates of discriminability and response bias that better reflect decision-making performance on an identification test. SDT-CD is specifically designed to describe performance for compound decisions (also known as simultaneous identification and detection decisions). Compound decisions comprise two components: detection and identification (Macmillan & Creelman, 1991). The detection component reflects the ability of a decision maker to detect the presence of a target stimulus in an array of stimuli. This might involve, for example, detecting a tumor in an x-ray. In the context of an identification test, it refers to witnesses’ ability to detect the presence of the culprit in a lineup. Perfect detection performance would involve choosing from all culprit-present lineups (regardless of whether the culprit or a foil is selected) and rejecting all culprit-absent lineups. In SDT-CD, detection performance is modeled as a 1-of-$m$ detection task, where $m$ is the number of lineup members.

The identification component (termed SDT-identification hereafter to distinguish it from the more general term “identification”) reflects a decision maker’s ability to attribute the correct identity to a stimulus. Because the SDT-identification component of compound decisions is distinct from the detection component, SDT-identification does not reflect the decision maker’s ability to detect the target; thus, this component of compound decisions is relevant only to positive responses made to arrays that contain a target stimulus. Determining the nature of a tumor that has been detected in an x-ray is an example of SDT-identification. In the context of eyewitness identification decisions, SDT-identification refers to witnesses’ ability to correctly identify the culprit, assuming that the culprit is in the lineup and the witness makes a positive identification. In SDT-CD, SDT-identification performance is modeled as an $m$-alternative, forced-choice decision.

In describing eyewitness identification performance, both the detection and SDT-components must be taken into account. SDT-CD does this by calculating response probabilities (i.e., foil identifications and correct lineup rejections for culprit-absent lineups, and correct identifications, foil identifications, and incorrect lineup rejections for culprit-present lineups) for all combinations of a single $d'$ value and single decision criterion value. (The estimated response bias index, $c$, is derived from the $d'$ and criterion values, with $c = \text{criterion} - d'/2$.) The model-generated response probabilities are compared against the observed response probabilities to find the combination of $d'$ and $c$ that best describes identification performance in the observed data. The better the model-generated response probabilities fit the observed response probabilities, the more appropriate it is to describe identification performance in terms of the relevant combination of $d'$ and $c$.

In this research, we used SDT-CD to reanalyze data from 22 published experiments that compared identification performance...
for sequential and simultaneous lineups. The aim was to investigate whether the difference in identification performance between sequential and simultaneous lineups could be characterized as a difference in discriminability, response bias, or a combination of both.

Method

Studies Included in the Analysis

The inclusion criteria were based on Steblay et al.’s (2011) criteria for a full diagnostic design. We included studies that (a) experimentally manipulated lineup presentation mode (sequential vs. simultaneous) and culprit-presence (culprit-present vs. culprit-absent), (b) produced above-chance identification performance (as defined by Steblay et al.), and (c) used adult participants (cf. children or older adults). For four studies that used multiple age groups (Memon & Gabbert, 2003; Parker & Ryan, 1993; Rose, Bull, & Vrij, 2005; Wilcock, Bull, & Vrij, 2005), the data for adults were included because they could be separated from the data for children and older participants.

We excluded data from three studies (Cutler & Penrod, 1988; Lindsay, Lea, Nosworthy, et al., 1991, Experiment 2; Lindsay, Pozzulo, Craig, Lee, & Corber, 1997, Experiment 1) because the response frequencies required for the computation of SDT-CD estimates of and (i.e., correct identifications and foil identifications for culprit-present lineups, and foil identifications for culprit-absent lineups) were not reported and could not be calculated from the reported data or obtained from the authors. We also excluded data from one study (Lindsay, Lea, Nosworthy, et al., 1991, Experiment 1) in which the manipulation of presentation mode was intentionally confounded with lineup instructions and foil similarity. Excluding this study had a negligible impact on the descriptive statistics and did not alter the results of any inferential tests.

Table 1 contains a list of the 22 studies included in the meta-analysis. Sample sizes ranged from 32 to 619 (mean [M] = 181.09, standard deviation [SD] = 124.54, total N = 3,984). Six-person lineups were used in all studies bar one (Lindsay, Lea, & Fulford, 1991) which used an eight-person lineup. In addition, Clark and Davey (2005) used a six-person culprit-present lineup and a five-person culprit-absent lineup.

Biased Versus Unbiased Lineups

We considered the role of lineup bias as a proposed moderator of the sequential lineup advantage (Carlson, Gronlund, & Clark, 2008). There is evidence that a sequential lineup advantage emerges only for biased lineups (Carlson et al., 2008), although some researchers have reported that this is not the case (e.g., Steblay et al., 2011). Of the studies included in the present meta-analysis, five were coded as using culprit-absent lineups that were biased toward the selection of an innocent suspect (Carlson et al., 2008, Experiment 2; Clark & Davey, Experiments 1 & 2; Lindsay & Wells, 1985; Lindsay, Lea, & Fulford, 1991). All other studies were coded as using unbiased lineups.

Statistical Procedures

The analyses focused on standardized effect sizes for the difference (in terms of and ) between sequential and simultaneous lineups. Several steps were involved in computing these standardized effect sizes.

First, we calculated SDT-CD estimates of and for each presentation condition (sequential and simultaneous) in each study. A broad range of possible parameter values was considered in the model selection process (i.e., ) for six-person lineups and for eight-person lineups). Following Palmer et al. (2010), we compared the observed and model-generated response probabilities using likelihood ratio G-tests (Sokal & Rohlf, 1981) in order to identify and assess the best-fitting combination of and values for each condition. These are conceptually the same as tests, but allow the G statistic to be summed across tests. For each condition, three separate G-tests were conducted to compare observed and model-generated expected frequencies of false alarms, hits (i.e., positive responses to culprit-present lineups), and correct identifications. (Expected frequencies were calculated by multiplying model-generated response probabilities by the number of target-absent and target-present trials for false alarms and hits, respectively, and by the number of observed hits in the case of correct identifications.) A statistic was calculated by summing G for the three tests. The smallest value indicated the best-fitting combination of and estimates for that condition.

Second, we used a modified jackknife procedure in order to obtain estimates of variance for each and value (Koriat, Lichtenstein, & Fischhoff, 1980; Mosteller & Tukey, 1968; Weber & Brewer, 2006). For each condition in each study, and were calculated as many times as there were participants, with a different participant omitted each time. For example, in a condition with 60 participants, and would be calculated 60 times, each time based on data from 59 participants. The resulting distributions of and values were used to calculate a jackknife estimate of the standard error and standard deviation for and in each condition.

For each study, we used the and values and associated jackknife estimates of standard deviation to compute two standardized effect size estimates for the difference between sequential and simultaneous lineup conditions (one for the difference in terms of and one for the difference in terms of ). The effect size index used was Hedge’s , which is conceptually the same as Cohen’s , but includes a correction to counter the overestimation of effect size in studies with small samples; Hedge’s would be calculated by summing all values consistent with the conclusions drawn in this paper. More information regarding this analysis is available from the authors.

1 We note that Steblay et al.’s (2011) meta-analysis leaves out one very large study (Gronlund et al., 2009) that was excluded because of below-chance identification performance. Analysis of that data set yielded the same conclusion drawn in this paper. More information regarding this analysis is available from the authors.

2 There are two versions of SDT-CD, each assuming that decision makers adopt a different decision rule (Duncan, 2006, Duncan, 2011). The integration rule assumes that decision-makers make a global assessment of the array, comparing the sum of strength-of-evidence values for all stimuli against the decision criterion. The independent observation rule assumes that each stimulus in the array is assessed separately against the decision criterion; a positive response is made if the strength of evidence for one or more stimuli exceeds the criterion. Here, we report only estimates based on the integration decision rule, as previous research indicates that this model provides a better fit for empirical data for eyewitness identification decisions (Duncan, 2006, Duncan, 2011; Palmer et al., 2010).
Table 1
Observed and Model-Generated Response Probabilities, Model Fit Statistic ($G_{total}$), and Estimates of Discriminability ($d'$) and Response Bias ($c$) for Each Study

<table>
<thead>
<tr>
<th>Study and condition</th>
<th>CID</th>
<th>FID</th>
<th>FA</th>
<th>CID</th>
<th>FID</th>
<th>FA</th>
<th>$G_{total}$</th>
<th>$d'$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlson et al. (2008, Exp. 1)</td>
<td>.72</td>
<td>.08</td>
<td>.23</td>
<td>.69</td>
<td>.05</td>
<td>.29</td>
<td>2.00</td>
<td>2.95</td>
<td>-0.91</td>
</tr>
<tr>
<td>Sequential</td>
<td>.57</td>
<td>.02</td>
<td>.16</td>
<td>.59</td>
<td>.02</td>
<td>.14</td>
<td>.21</td>
<td>3.29</td>
<td>-0.58</td>
</tr>
<tr>
<td>Carlson et al. (2008, Exp. 2)</td>
<td>.48</td>
<td>.17</td>
<td>.65</td>
<td>.50</td>
<td>.28</td>
<td>.54</td>
<td>22.59</td>
<td>1.58</td>
<td>-0.90</td>
</tr>
<tr>
<td>Sequential</td>
<td>.37</td>
<td>.15</td>
<td>.44</td>
<td>.38</td>
<td>.22</td>
<td>.35</td>
<td>10.62</td>
<td>1.57</td>
<td>-0.40</td>
</tr>
<tr>
<td>Clark &amp; Davey (2005, Exp. 1)</td>
<td>.25</td>
<td>.54</td>
<td>.73</td>
<td>.25</td>
<td>.55</td>
<td>.73</td>
<td>0.01</td>
<td>0.57</td>
<td>-0.89</td>
</tr>
<tr>
<td>Sequential</td>
<td>.46</td>
<td>.40</td>
<td>.71</td>
<td>.46</td>
<td>.40</td>
<td>.71</td>
<td>0.00</td>
<td>1.26</td>
<td>-1.18</td>
</tr>
<tr>
<td>Clark &amp; Davey (2005, Exp. 2)</td>
<td>.42</td>
<td>.44</td>
<td>.67</td>
<td>.42</td>
<td>.41</td>
<td>.69</td>
<td>0.31</td>
<td>1.16</td>
<td>-1.07</td>
</tr>
<tr>
<td>Sequential</td>
<td>.48</td>
<td>.33</td>
<td>.65</td>
<td>.48</td>
<td>.34</td>
<td>.64</td>
<td>0.05</td>
<td>1.40</td>
<td>-1.05</td>
</tr>
<tr>
<td>Greathouse &amp; Kovera (2009)</td>
<td>.60</td>
<td>.28</td>
<td>.90</td>
<td>.60</td>
<td>.34</td>
<td>.83</td>
<td>6.53</td>
<td>1.57</td>
<td>-1.74</td>
</tr>
<tr>
<td>Sequential</td>
<td>.60</td>
<td>.33</td>
<td>.83</td>
<td>.60</td>
<td>.34</td>
<td>.82</td>
<td>0.16</td>
<td>1.59</td>
<td>-1.70</td>
</tr>
<tr>
<td>Kneller et al. (2001)</td>
<td>.61</td>
<td>.17</td>
<td>.61</td>
<td>.62</td>
<td>.20</td>
<td>.56</td>
<td>0.61</td>
<td>1.98</td>
<td>-1.13</td>
</tr>
<tr>
<td>Sequential</td>
<td>.50</td>
<td>.11</td>
<td>.22</td>
<td>.49</td>
<td>.10</td>
<td>.24</td>
<td>0.07</td>
<td>2.32</td>
<td>-0.45</td>
</tr>
<tr>
<td>Levi (2006)</td>
<td>.63</td>
<td>.23</td>
<td>.49</td>
<td>.62</td>
<td>.20</td>
<td>.53</td>
<td>0.74</td>
<td>2.01</td>
<td>-1.08</td>
</tr>
<tr>
<td>Sequential</td>
<td>.35</td>
<td>.18</td>
<td>.30</td>
<td>.35</td>
<td>.19</td>
<td>.28</td>
<td>0.09</td>
<td>1.64</td>
<td>-0.25</td>
</tr>
<tr>
<td>Lindsay, Lea, &amp; Fulford (1991)</td>
<td>.57</td>
<td>.20</td>
<td>.57</td>
<td>.57</td>
<td>.23</td>
<td>.55</td>
<td>0.31</td>
<td>2.02</td>
<td>-1.13</td>
</tr>
<tr>
<td>Sequential</td>
<td>.47</td>
<td>.07</td>
<td>.17</td>
<td>.46</td>
<td>.06</td>
<td>.17</td>
<td>0.07</td>
<td>2.79</td>
<td>-0.46</td>
</tr>
<tr>
<td>Lindsay &amp; Wells (1985)</td>
<td>.58</td>
<td>.12</td>
<td>.58</td>
<td>.61</td>
<td>.18</td>
<td>.49</td>
<td>5.89</td>
<td>2.04</td>
<td>-0.99</td>
</tr>
<tr>
<td>Sequential</td>
<td>.50</td>
<td>.02</td>
<td>.35</td>
<td>.55</td>
<td>.08</td>
<td>.24</td>
<td>10.21</td>
<td>2.51</td>
<td>-0.56</td>
</tr>
<tr>
<td>MacLin &amp; Phelan (2007)</td>
<td>.48</td>
<td>.25</td>
<td>.50</td>
<td>.48</td>
<td>.26</td>
<td>.49</td>
<td>0.04</td>
<td>1.63</td>
<td>-0.79</td>
</tr>
<tr>
<td>Sequential</td>
<td>.23</td>
<td>.08</td>
<td>.10</td>
<td>.23</td>
<td>.08</td>
<td>.10</td>
<td>0.01</td>
<td>1.95</td>
<td>0.33</td>
</tr>
<tr>
<td>MacLin et al. (2005, Exp. 1)</td>
<td>.40</td>
<td>.43</td>
<td>.63</td>
<td>.40</td>
<td>.41</td>
<td>.66</td>
<td>0.23</td>
<td>1.13</td>
<td>-0.98</td>
</tr>
<tr>
<td>Sequential</td>
<td>.33</td>
<td>.17</td>
<td>.40</td>
<td>.35</td>
<td>.22</td>
<td>.33</td>
<td>1.39</td>
<td>1.48</td>
<td>-0.31</td>
</tr>
<tr>
<td>MacLin et al. (2005, Exp. 2)</td>
<td>.47</td>
<td>.23</td>
<td>.50</td>
<td>.47</td>
<td>.26</td>
<td>.48</td>
<td>0.19</td>
<td>1.61</td>
<td>-0.75</td>
</tr>
<tr>
<td>Sequential</td>
<td>.27</td>
<td>.17</td>
<td>.23</td>
<td>.27</td>
<td>.17</td>
<td>.23</td>
<td>0.02</td>
<td>1.48</td>
<td>0.01</td>
</tr>
<tr>
<td>Melara et al. (1989)</td>
<td>.25</td>
<td>.63</td>
<td>.94</td>
<td>.24</td>
<td>.68</td>
<td>.89</td>
<td>0.40</td>
<td>0.38</td>
<td>-1.44</td>
</tr>
<tr>
<td>Sequential</td>
<td>.13</td>
<td>.25</td>
<td>.25</td>
<td>.13</td>
<td>.24</td>
<td>.26</td>
<td>0.02</td>
<td>0.68</td>
<td>0.29</td>
</tr>
<tr>
<td>Memon &amp; Gabbert (2003)</td>
<td>.47</td>
<td>.33</td>
<td>.53</td>
<td>.47</td>
<td>.31</td>
<td>.56</td>
<td>0.25</td>
<td>1.47</td>
<td>-0.89</td>
</tr>
<tr>
<td>Sequential</td>
<td>.17</td>
<td>.07</td>
<td>.10</td>
<td>.17</td>
<td>.08</td>
<td>.09</td>
<td>0.18</td>
<td>1.68</td>
<td>0.51</td>
</tr>
<tr>
<td>Parker &amp; Ryan (1993)</td>
<td>.38</td>
<td>.17</td>
<td>.42</td>
<td>.39</td>
<td>.22</td>
<td>.36</td>
<td>0.98</td>
<td>1.57</td>
<td>-0.41</td>
</tr>
<tr>
<td>Sequential</td>
<td>.29</td>
<td>.25</td>
<td>.33</td>
<td>.29</td>
<td>.25</td>
<td>.34</td>
<td>0.00</td>
<td>1.27</td>
<td>-0.21</td>
</tr>
<tr>
<td>Pozzulo et al. (2008)</td>
<td>.48</td>
<td>.28</td>
<td>.53</td>
<td>.48</td>
<td>.28</td>
<td>.53</td>
<td>0.00</td>
<td>1.55</td>
<td>-0.86</td>
</tr>
<tr>
<td>Sequential</td>
<td>.40</td>
<td>.16</td>
<td>.24</td>
<td>.39</td>
<td>.15</td>
<td>.25</td>
<td>0.07</td>
<td>1.87</td>
<td>-0.28</td>
</tr>
<tr>
<td>Pozzulo &amp; Marciniak (2006)</td>
<td>.45</td>
<td>.25</td>
<td>.53</td>
<td>.45</td>
<td>.28</td>
<td>.50</td>
<td>0.68</td>
<td>1.51</td>
<td>-0.76</td>
</tr>
<tr>
<td>Sequential</td>
<td>.35</td>
<td>.57</td>
<td>.72</td>
<td>.36</td>
<td>.47</td>
<td>.71</td>
<td>7.04</td>
<td>0.95</td>
<td>-1.03</td>
</tr>
<tr>
<td>Rose et al. (2005)</td>
<td>.75</td>
<td>.21</td>
<td>.38</td>
<td>.71</td>
<td>.13</td>
<td>.51</td>
<td>5.71</td>
<td>2.38</td>
<td>-1.21</td>
</tr>
<tr>
<td>Sequential</td>
<td>.46</td>
<td>.17</td>
<td>.29</td>
<td>.46</td>
<td>.16</td>
<td>.31</td>
<td>0.05</td>
<td>1.95</td>
<td>-0.47</td>
</tr>
<tr>
<td>Sporer (1993)</td>
<td>.44</td>
<td>.33</td>
<td>.72</td>
<td>.44</td>
<td>.38</td>
<td>.67</td>
<td>0.62</td>
<td>1.26</td>
<td>-1.06</td>
</tr>
<tr>
<td>Sequential</td>
<td>.39</td>
<td>.17</td>
<td>.39</td>
<td>.40</td>
<td>.21</td>
<td>.34</td>
<td>0.43</td>
<td>1.64</td>
<td>-0.41</td>
</tr>
<tr>
<td>Steblay et al. (in press)</td>
<td>.25</td>
<td>.17</td>
<td>.27</td>
<td>.25</td>
<td>.19</td>
<td>.25</td>
<td>0.72</td>
<td>1.33</td>
<td>0.02</td>
</tr>
<tr>
<td>Sequential</td>
<td>.26</td>
<td>.20</td>
<td>.26</td>
<td>.26</td>
<td>.20</td>
<td>.25</td>
<td>0.00</td>
<td>1.35</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

(table continued)
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Study and condition</th>
<th>Observed</th>
<th></th>
<th></th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CID</td>
<td>FID</td>
<td>FA</td>
<td>$G_{total}$</td>
</tr>
<tr>
<td>Wells &amp; Pozzulo (2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simultaneous</td>
<td>.32</td>
<td>.32</td>
<td>.50</td>
<td>.32</td>
</tr>
<tr>
<td>Sequential</td>
<td>.16</td>
<td>.40</td>
<td>.48</td>
<td>.16</td>
</tr>
<tr>
<td>Wilcock et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simultaneous</td>
<td>.67</td>
<td>.13</td>
<td>.63</td>
<td>.68</td>
</tr>
<tr>
<td>Sequential</td>
<td>.63</td>
<td>.04</td>
<td>.17</td>
<td>.62</td>
</tr>
</tbody>
</table>

Note. CID = correct identifications from culprit-present lineups; FID = foil identifications from culprit-present lineups; FA = foil identifications from culprit-absent lineups (including suspect and foil identifications).

[3/(4N − 9)]. Because most of the studies included had relatively large samples, this correction had a negligible impact on effect size estimates; it reduced one estimate by 0.025 (for the $c$ value in Melara, DeWitt-Rickards, & O’Brien, 1989; $N = 32$) and all others by less than 0.006.

The effect sizes for individual studies were combined (Rosenthal, 1991) to produce weighted mean effect sizes (one for $d’$ and one for $c$ values), and associated $z$-scores and probability values. These represented the difference between sequential and simultaneous lineups across all of the included studies. For significant results, a failsafe $N (N_{fs})$ was calculated, representing the number of “file drawer” studies that would be required to increase the $p$ value to .05 (Rosenthal).

Results

Model Fit

Table 1 shows the observed and model-generated response probabilities, $G_{total}$ statistics, and estimates of $d’$ and $c$ for the simultaneous and sequential lineup conditions in each study. The best-fitting SDT-CD model provided a good fit to the data for the majority of conditions ($G_{total} < 1$ in 34 of 44 conditions). The most notable exceptions came from two studies in which the culprit-absent lineup contained an innocent suspect chosen to be a better match than other lineup members to the culprit (Carlson et al., 2008, Experiment 2; Lindsay & Wells, 1985). The relatively poor fit for these studies was not surprising, given that SDT-CD makes the assumption that all lineup foils are equally likely to be chosen from the lineup. SDT-CD has difficulty reconciling the response probabilities for culprit-present lineups with the relatively high foil identification rate typically found for biased culprit-absent lineups (Luus & Wells, 1991). For example, the model that best fits the culprit-present data for simultaneous lineups in Carlson et al. (2008, Experiment 2) predicts that the foil identification rate for culprit-absent lineups will be .36, which is much lower than the observed rate of .65.

Aside from these studies, the $G_{total}$ values were generally low and the model-generated response probabilities closely aligned with the data, indicating that identification performance could reasonably be described in terms of the model-generated estimates of $d’$ and $c$.

The Effects of Presentation Mode on Estimates of $d’$ and $c$

Hedge’s $d$ values can be interpreted as per Cohen’s $d$, with cutoffs for small, medium, and large effects of 0.2, 0.5, and 0.8, respectively. For comparisons of $d’$, positive values for Hedge’s $d$ indicate greater discriminability for sequential (cf. simultaneous) lineups. For comparisons of $c$, positive values indicate more conservative bias for sequential (cf. simultaneous) lineups.

The results were clear cut, and best summarized in terms of the weighted mean effect sizes for $d’$ and $c$ and the associated $z$ scores. For $d’$ estimates, the weighted mean effect size was negligible and nonsignificant; mean Hedge’s $d = 0.04$, standard error [SE] = 0.03, CI$_{95}$ (−0.03, 0.10), $z = 1.13, p = .13$. Thus, $d’$ did not differ between sequential and simultaneous lineups. This finding was also evident in the pattern of effect sizes for individual studies. As shown in the top panel of Figure 2, only three studies met the benchmark of 0.2 for a small effect (with another just below this cutoff), and the standard errors overlapped a Hedge’s $d$ value of zero for the majority of studies.

For $c$ values, the weighted mean effect size was small, significant, and positive; mean $c = 0.23, SE = 0.03, CI_{95} (0.17, 0.30), z = 7.29, p < .001, N_{total} = 494$. Thus, responding was more conservative for sequential than simultaneous lineups. The pattern of effect sizes for individual studies supported this conclusion. As shown in the lower panel of Figure 2, for the majority of studies, the effect size for the comparison of $c$ values exceeded the cutoff for a small effect and the standard errors did not overlap zero.

It is important that the results show that sequential (cf. simultaneous) presentation produced not only more conservative responding, but also less-biased responding. A $c$ value of zero represents unbiased responding, negative values represent lenient responding, and positive values represent conservative responding. Although responding was markedly lenient for both presentation modes, witnesses—on average—set a less-biased response criterion for sequential lineups (unweighted mean $c = −0.41, SE = 0.01$) than simultaneous lineups (unweighted mean $c = −0.94, SE = 0.01$). At the individual study level, responding was less biased for the sequential (cf. simultaneous) condition in 19 of the 22 studies analyzed (see Table 1).

Biased versus unbiased lineups. Lineup bias did not moderate the effects of presentation mode on estimates of $d’$, $Q_{d’}(1) = 2.05, p = .15$, or $c$, $Q_{c}(1) < 1$. ($Q_{d’}$ is a statistic that reflects between-group homogeneity and follows a $χ^2$ distribution.) How-
ever, when the $d'$ values were examined separately for biased and unbiased lineups, the pattern of weighted mean effects sizes did offer a hint of support for Carlson et al.'s (2008) conclusions: there was a trend toward slightly better discriminability for sequential (cf. simultaneous) presentation for studies that used biased lineups (mean Hedge's $d = 0.11$, $SE = 0.05$) but not those that used unbiased lineups (mean Hedge's $d = 0.01$, $SE = 0.04$).

**Discussion**

This research provides the first direct evidence that the sequential lineup advantage arises because, as compared with simultaneous presentation, sequential presentation prompts witnesses to adopt a decision criterion that is not only more conservative, but also less biased. This finding has some important implications. From a theoretical perspective, the results extend knowledge about the eyewitness identification decision process. For example, this research informs understanding of Wells' (1984) conceptual distinction between *relative judgments* (whereby witnesses compare lineup members and select the person who most closely resembles their memory of the culprit) and *absolute judgments* (whereby witnesses compare each lineup member to their memory of the culprit and only make a positive identification if there is a sufficiently good match). The development of the sequential lineup procedure was based on the relative–absolute conceptualization; Lindsay and Wells (1985) reasoned that sequential presentation of lineup members would encourage witnesses to make absolute rather than relative judgments. These results do not directly contradict the relative–absolute account of the sequential lineup advantage. They do, however, have implications for how we view the difference between relative and absolute judgments. To the extent that the sequential lineup advantage is due to a shift from relative to absolute judgments (Gronlund, 2004; Lindsay et al., 2009; Lindsay & Wells, 1985), the relative–absolute distinction represents—in terms of performance—a difference in response bias.

In addition, the results of this research have the potential to advance theory by shaping the development of formal models of the eyewitness identification decision process (e.g., WITNESS; Clark, 2003). The application of SDT-CD to eyewitness identification data does not shed light directly on the cognitive processes that underpin identification decisions. Instead, the value of SDT-CD lies in enabling identification performance to be described in terms of discriminability and response bias which, in turn, can inform the development of models that do explore the cognitive processes involved in identification decisions. The findings of the present research indicate that models of identification decisions must include some mechanism for accounting for the difference in performance between sequential and simultaneous lineups in terms of a shift in response criterion (Goodsell et al., 2010).

From an applied viewpoint, this research sheds light on one of the key points in the ongoing debate about whether researchers should endorse sequential lineups as a replacement for simultaneous lineups (e.g., Malpass et al., 2009; McQuiston–Surrett et al., 2006; Meissner et al., 2005). On one hand, the results demonstrate that it is quite appropriate to claim that the sequential advantage is due to a conservative shift in responding rather than to improved discriminability. However, on the other, the results indicate that sequential (cf. simultaneous) presentation reduces bias in criterion setting. Therefore, the fact that the sequential advantage reflects a conservative shift in responding does not necessarily constitute a weakness of the sequential lineup procedure; rather, it is this conservative shift in responding that is responsible for the higher accuracy rates observed for sequential (cf. simultaneous) presentation in experimental settings. (This finding is not inconsistent with the view that sequential presentation reduces guessing in identification responses; Penrod, 2003.)

It is important to note that although this study shows that sequential presentation promotes less-biased criterion setting, it does not necessarily follow that sequential lineup presentation will be superior to simultaneous presentation in real world settings. As mentioned earlier, less-biased responding does not always produce higher accuracy rates. For example, in situations where culprit-present lineups are far more common than culprit-absent lineups, accuracy is maximized when responding is lenient, rather than unbiased. Conversely, when the base rate of culprit-present lineups

![Figure 2. Effect sizes for comparisons of $d'$ and $c$ for each of the 22 studies analyzed, ordered by magnitude of effect. Positive values indicate greater discriminability (for $d'$) and more conservative responding (for $c$) for sequential versus simultaneous lineups. Cutoffs for small, medium, and large effects are 0.2, 0.5, and 0.8, respectively. Error bars denote standard errors.](image-url)
is low, accuracy is maximized by conservative responding. Because the base rate of culprit-present lineups in actual police investigations is not known, we cannot be certain whether less-biased responding will produce greater accuracy in these settings. Further, our research does not speak to the problem of how to weigh the value of a potential wrongful conviction against a potential wrongful acquittal (Malpass, 2006; Steblay et al., 2011; Volokh, 1997). Based on the idea that the protection of the innocent should be prioritized over seeking justice for the guilty (e.g., Laufer, 1995), a lineup procedure that promotes more conservative choosing might be favored. However, this is an extremely complex issue that has occupied legal professionals and philosophers for centuries, and remains unresolved in many jurisdictions (see Volokh for a review).

Thus, although the present research advances understanding of the effect of presentation mode on eyewitness identification decisions, it does not address all of the issues that have been raised in the sequential versus simultaneous debate. Indeed, perhaps the most striking ideas to emerge from the debate are (a) that many substantial problems associated with identification tests have nothing to do with presentation mode (e.g., the use of biased instructions and nondouble-blinding; Wells, 2001), and (b) neither sequential nor simultaneous lineups produce adequately high accuracy rates (Lindsay et al., 2009). Thus, as well as continuing to evaluate established lineup methods, researchers would do well to build on current knowledge in order to develop new, more effective techniques for conducting identification tests (Wells, Memon, & Penrod, 2006). One example of such a test involves the use of multiple lineups that assess different aspects of the witness’s memory for the culprit, such as the face, body, clothing, and voice (Pryke, Lindsay, Dysart, & Dupuis, 2004). In another line of research, rather than making a traditional binary lineup response, the witness is asked to provide a rating for each lineup member that indicates how confident they are that person is the culprit (Sauer, Brewer, & Weber, 2008). And, in the elimination lineup procedure (Pozzulo & Lindsay, 1999), identification decisions are separated into two stages by asking witnesses to first select the lineup member that best matches their memory of the culprit and then decide whether that person is the culprit. Although these two separate decisions do not map directly on to the detection and SDT-identification components of a compound decision, approaches that take into account the relatively complex nature of lineup responses may prove fruitful.

References

References marked with an asterisk indicate studies included in the meta-analysis.


Lindsay, R. C. L. & Wells, G. L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presenta-


Received March 11, 2011
Revision received June 8, 2011
Accepted June 11, 2011