Loughborough University

This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Simultaneous v. sequential lineups: What do we really know?

PLEASE CITE THE PUBLISHED VERSION

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Ebbesen, Ebbe B., and Heather D. Flowe. 2019. “Simultaneous V. Sequential Lineups: What Do We Really Know?”. figshare. https://hdl.handle.net/2134/20167.

Simultaneous v. Sequential Lineups: What Do We Really Know?

Ebbe B. Ebbesen and Heather D. Flowe

Department of Psychology

University of California, San Diego

DRAFT

Author Note

Ebbe B. Ebbesen and Heather D. Flowe, Department of Psychology, University of California, San Diego.

This work is funded by U.S. Grants and supported under a National Science Foundation Graduate Research Fellowship to the second author. A list of all references used for the meta-analysis is available at: http:www-psy.ucsd.edu/~eebbesen/SimSeq.htm. Address correspondence to Ebbe B. Ebbesen, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92090-0109; email: eebbesen@ucsd.edu.

Abstract

Both conceptual and meta-analyses of the effects of simultaneous and sequential lineup testing procedures on false alarm and hit rates suggest that recent interest in moving to sequential lineups might be premature. A simple criterion-shift model based on signal detection analysis accounted for the results from the meta-analysis raising concern that the previously accepted relative v. absolute decision strategy view may be incorrect. The accepted view that hit rates will be unaffected by a change in procedure may be incorrect. Monte Carlo simulation results raise the possibility that serial position might play a much larger and more complicated role in performance on sequential lineups than has been considered. Much more research is needed before the sequential procedure is adopted.

DRAFT

INTRODUCTION

Recently, the National Institute of Justice collected a group of detectives, prosecutors, defense attorneys, and psychologists together to write guidelines for various agencies in the criminal justice system. The effort resulted in (Eyewitness evidence: A guide for law enforcement, 1999). This document provides procedures for improving the collection and documentation of eyewitness testimony and identification, especially recommendations for conducting lineups. In particular, the guidelines suggest that sequential lineups produce more reliable identifications than simultaneous lineups**.** In simultaneous lineups, investigators present all of the potential choices together in time. In sequential lineups, the choices are presented one at a time in sequence. Studies comparing the two procedures have generally concluded that the rate of false alarms is lower for sequential than simultaneous lineups while the rate of hits is no different. Although the guidelines stop short of explicitly recommending one lineup procedure over another (due to a lack of consensus on the issue), police agencies could change their practice based on the guideline's assertion that sequential lineups produce higher accuracy rates. For example, New Jersey, the first state in the US to incorporate the guidelines into police officer training, urges its officers to conduct sequential lineups when possible (http://www.state.nj.us/lps/dcj/agguide/photoid.pdf).

Adopting one lineup procedure over another obviously could have a significant impact on criminal case outcomes (Levi, 1998). As such, it is important to evaluate the decision strategy models and empirical evidence that have been advanced to account for the differential accuracy rates observed between the two lineup procedures. The assumed advantage of the sequential procedure is based on three things: 1) models of the purportedly different decision strategies that arise from the two test procedures (Lindsay & Wells, 1985), 2) evidence from laboratory research that seems consistent with the proposed model of how witnesses choose from simultaneous lineups (Lindsay, Lea, Nosworthy, & Fulford, 1991; Wells, 1993; Wells et al., 1998), and 3) laboratory research comparing accuracy rates across the two lineup procedures (Cutler & Penrod, 1988; Levi, 1998; Lindsay, Lea, & Fulford, 1991; Lindsay, Lea, Nosworthy et al., 1991; Lindsay &

DRAFT

Wells, 1985; Lindsay, 1999; Lindsay, Pozzulo, Craig, & Lee, 1997; Parker & Ryan, 1993; Sporer, 1994).

In this article, we review the empirical evidence and conceptually analyze the decision models that have been developed to account for the differential accuracy rates observed in simultaneous and sequential lineup studies. Signal detection theory and Monte-Carlo simulations of eyewitness decision-making in the two lineup procedures aid our analysis. We also present a meta-analysis of the accuracy rates obtained in the past 25 years of published lineup identification research. Overall, the analyses suggest that an alternative view, a criterion-shift model based on signal detection theory, better accounts for apparent differences in identification accuracy across the two procedures. This account raises as yet unresolved issues that require considerably more research before we fully understand eyewitness performance in lineups. We conclude that it seems premature to recommend the universal adoption of the sequential over the simultaneous lineup procedure at this time.

RESEARCH AND THEORY ON THE DECISION MODELS

A widely accepted explanation for the differences between simultaneous and sequential lineups is that eyewitnesses use a relative decision strategy when examining a simultaneous lineup and an absolute decision strategy when looking at a sequentially presented one (Lindsay, Lea, Nosworthy et al., 1991; Lindsay & Wells, 1985; Lindsay, 1999; Lindsay et al., 1997; Sporer, 1994; Wells et al., 1998).[1] The logic is that when multiple faces are presented simultaneously, witnesses will compare all of the choices to each other looking for the most familiar person in the display. When the most familiar face is selected, they will choose that person as the culprit. In contrast, when faces are presented one at a time sequentially, an absolute "identity" judgment is made because the opportunity to compare across the alternatives is considerably reduced. Witnesses are presumably forced to base their decision on how well each person's appearance matches (or is inconsistent with) information stored in memory about the culprit. Consequently, the most familiar person will be chosen only if that person sufficiently matches the

[1] This view has also been used to explain the difference in eyewitness performance when eyewitnesses are tested with a simultaneous lineup compared to a show-up (one and only one choice -- "Is that him?") procedure (Behrman & Vayder, 1994; Gonzalez, Ellsworth, & Pembroke, 1993; Lindsay et al., 1997; Yarmey, 1998; Yarmey, Yarmey, & Yarmey, 1996).

DRAFT

contents of memory.[2] If no one matches memory to a sufficient extent, the witness rejects the entire lineup.

Laboratory studies suggest that the use of the sequential procedure can lower the rate at which innocent suspects are chosen from culprit absent lineups. Additionally, these studies seem to find that the rate at which the actual culprit is chosen when present in the lineup does not differ between simultaneous and sequential lineups. As described, this outcome protects the innocent without reducing the odds that the guilty will go free—clearly a desirable outcome.

Empirical Evidence for Relative Judgment in Simultaneous Lineups

Some might argue that sufficient experimental evidence already exists to support the claim that eyewitnesses use a relative judgment strategy in simultaneous lineups and an absolute one in sequential lineups. In fact, New Jersey Attorney General John Farmer, in his guidelines calling for the use of sequential lineups, reported that scientific studies have "proven that witnesses have a tendency to compare one member of a lineup to another, making relative judgments about which individual looks most like the perpetrator" (http://www.state.nj.us/lps/dcj/agguide/photoid.pdf). Despite the Attorney General's claim, the main concern of research examining different lineup procedures has been whether the two lineup procedures produce different accuracy rates and not the precise nature of the differences in decision strategy in the two procedures. Moreover, no experiment to date has been designed to test the idea that eyewitnesses use an absolute judgment strategy when confronted with a sequential lineup. In this section, research findings that are usually cited as evidence for relative judgment in simultaneous lineups (Wells et al., 1998) are reviewed. These include research involving the removal without replacement procedure, lineup admonishment, self-report data on judgment strategy, the dual lineup procedure, and lineup member to perpetrator similarity. Throughout, we argue that the research, hardly any of which was originally designed to test the

[2] The strength and content of witness memories should be the same regardless of the lineup procedure used to assess those memories. Procedural differences between simultaneous and sequential lineup presentation, however, might affect "retrieval" or "reconstructive" processes in such a manner that the information extracted from memory will be different for the two procedures.

DRAFT

parameters of the relative judgment model, is far from conclusive, in part because it fails to eliminate the alternative explanation that witnesses employ absolute comparisons in both simultaneous and sequential lineups.

Throughout this paper, we assume that identification (or recognition) is inherently a comparison task. That is, unless witnesses are just guessing, they would be expected to compare the items presented to them with some representation of information that they have in memory, even if that information is familiarity-based. The details of what is compared and exactly how the comparison is accomplished need not be made specific to understand that such comparisons might result in a wide range of values. Some items would yield very good matches and some very poor matches depending on what is in memory and the nature of the items that are presented to the witnesses. If this view is correct, witnesses would have to decide whether the degree of match was sufficient to claim identity. That is, a witness would have to set a criterion for deciding whether presented faces were sufficiently identical to the contents of his or her memory of the culprit before claiming that the face and the memory are the same. We argue that the possibility that witnesses might set "absolute" degree-of-match criteria in both sequential and simultaneous lineups forms the basis for an alternative explanation of findings that have been taken as support for the relative decision model.

Removal without replacement

The "removal without replacement" procedure (Wells, 1993) has been cited as the best evidence that eyewitnesses use a relative judgment strategy when confronted with a simultaneous lineup (Wells et al., 1998). The removal without replacement procedure involves two parts. After viewing a staged crime in the laboratory, eyewitnesses are shown a culprit-present simultaneous lineup. Both the distribution of choices over all available alternatives in the lineup and the no choice rate for this group are recorded. In the second part, another group of eyewitnesses undergoes similar experimental procedures, except that the target's picture has been removed from the lineup and not replaced with any other photograph. According to Wells (1993), if a relative judgment strategy is being used, the eyewitnesses viewing the target-removed lineup will be inclined to select the most familiar face from the remaining set of foils instead of

DRAFT

correctly rejecting the lineup. This prediction assumes that the choices of the target in the target present lineup were based on the fact that the target was the most familiar of all of the presented alternatives. On the other hand, if witnesses were using an absolute judgment strategy in such lineups, then when viewing the target-removed lineup, they should correctly reject the lineup because the culprit is not there. As shown in Figure 1, (Wells, 1993) found that the subjects were more likely to pick a foil and less likely to reject the lineup after the target had been removed.

Insert Figure 1

On the other hand, the relative decision strategy model assumes that witnesses examine all of the choices in a simultaneous lineup and then choose the most familiar alternative. Were this the strategy, witnesses would always choose someone from simultaneous lineups because, except in the rare case of a tie, one alternative should always be the most familiar (regardless of how unfamiliar that alternative might be). Not only is the correct rejection rate for both groups in the Wells study higher than one might expect if all subjects were using a relative decision strategy, we know that witnesses frequently fail to choose someone from simultaneous lineups (even when the culprit is present) both in the real world (Behrman & Davey, 1999; Tollerstrup, Turtle, & Yuille, 1994) and in laboratory simulation studies (Lindsay, Lea, & Fulford, 1991; Lindsay & Wells, 1985; Lindsay et al., 1997; Pozzulo & Lindsay, 1999; Sporer, 1994; Yarmey & Morris, 1998). For example, in our meta-analysis (reported later in this paper), adult witnesses tested in 114 experiments with simultaneous target present lineups failed to pick anyone an average of 51.7% of the time. In 84 experiments employing target absent simultaneous lineups, the average correct rejection rate was 49.9%. In addition, about a third of the witnesses who viewed the target-removed lineup in the Wells (1993) study refused to pick the most familiar person when they rejected the entire lineup. In short, a refinement of the relative decision-strategy idea that witnesses always pick the most familiar person is therefore necessary to explain the fact that about half (taking all of the studies together) of all adult witnesses in published experiments fail to choose the most familiar alternative when the target is absent.

DRAFT

One possibility originally proposed by Lindsay & Wells (1985) and recently discussed by Gonzalez, Ellsworth, & Pembroke (1993) in the context of comparing single-suspect "showup" procedures to simultaneous lineups is the view that witnesses employ a combination of a relative and an absolute strategy in a simultaneous lineup.[3] Namely, they might look at all of the choices, pick the most familiar, and then compare this most familiar-looking person to the same kind of absolute similarity standard that is assumed to be used in showups and in sequential lineups.[4] That is, they might decide whether the most familiar person's looks matched those of the recalled culprit to a sufficient degree. If the match exceeds the absolute standard, then the witness would select that person as the culprit.

Once one allows for the possibility that witnesses confronted with a simultaneous lineup might do more than always select the most familiar option, it is possible to refine the relative decision model proposed to explain Wells' (1993) findings. In particular, it seems reasonable that more than one face will exceed the absolute criterion set by witnesses for some witness-lineup combinations. If so, witnesses would have to choose among the subset of presented faces that were familiar enough to exceed their criteria. Presumably, they would choose the face that exceeded the standard the most. If witnesses used this strategy in simultaneous lineups, they would be expected to behave as they did in Wells' (1993) removal without replacement study. After all, removing the target would mean that the one face most likely to exceed the standard was no longer available. However, one or more of the remaining faces, although less familiar than the target, might still be above the criterion set by witnesses. As a result, when presented with the remaining foils, some proportion of the witnesses would be expected to select one of them. How many witnesses would select one of the remaining foils would depend on the rate, over witnesses, at which one or more of the remaining foils exceeded whatever standard the witnesses set. Lower standards and greater perceived similarity between memory of the culprit and the foils would mean more witnesses from the second group would choose one of the remaining foils.

[3] The idea that both relative and absolute processes might be involved raises the interesting but empirically untested issue concerning whether need for multiple processes arises because different people use different strategies or because most individuals use both.

DRAFT

If one accepts the notion that witnesses might employ a criterion in simultaneous lineups, then what is different between simultaneous and sequential lineups? Faces are selected in both cases only if they exceed some absolute standard or "match criterion". One possible answer is that witnesses employ an absolute standard in both lineup procedures but they set the standard lower in simultaneous than sequential lineups. A lower standard in simultaneous lineups could explain why witnesses make more false alarms in target absent simultaneous than sequential lineups. On the other hand, this differential criterion view also predicts that witnesses should make more correct choices in simultaneous compared to the sequential procedure when the target is present. That is, a lower criterion would increase the rate at which witnesses chose both innocent "look-alikes" in blank lineups and the actual culprits in target present lineups because faces that were less than perfectly matched to memory representations would be more likely to exceed the lower criterion. Another way of stating this prediction is that witnesses should be more likely to make a choice with lower than higher decision standards. These predictions seem inconsistent with conclusions from individual studies that typically report that the simultaneous/sequential difference is in false alarm rates, but not in hit rates (Lindsay & Wells, 1985; Sporer, 1993). As a result, at first look, the idea that differentially strict absolute criteria are used in both types of lineups seems inconsistent with current findings.[5] We will return to this issue in subsequent sections in which we provide a signal detection analysis and the results from a meta-analysis of the lineup literature.

Effect of admonishment

A second source of evidence that relative judgments are made in simultaneous lineups is the effect that biased lineup instructions have on identification decisions (Wells et al., 1998). Eyewitnesses who are admonished that the culprit "may or may not be present" prior to viewing a target absent lineup are less apt to make false identifications

[4] Presumably were the absolute decision to result in a rejection, the witness would stop the process and say that they could not find the culprit among the presented alternatives.

[5] On the other hand, if it is true that the false alarm rate and not hit rate is affected by the change from simultaneous to sequential and the absolute standard that is used is set relatively high, then this means that the process by which witnesses select someone in target present lineups must be different than the process they use in target absent lineups. If the processes were the same, then both hit and false alarm rates should be affected by the change from simultaneous to sequential lineups.

DRAFT

than eyewitnesses who are not given such a suggestion. In a recent meta-analysis, Steblay (1997) found admonishing witnesses that the culprit might not be in the lineup increases the rate at which target absent lineups are rejected while having "minimal" effect on the rate at which the culprit is correctly identified. That is, admonishment to use a stricter criterion only seems to affect errors made to target absent lineups. One interpretation of this outcome is that admonishment discourages eyewitnesses from making relative judgments (Wells, et al. 1998). Presumably, if witnesses are led to believe that the culprit is in the lineup, they compare lineup choices and select the most familiar alternative. Witnesses who do not possess such a belief avoid making familiarity judgments and instead compare each lineup member to their actual memories of the culprit.

Though this is a plausible account, multiple alternative hypotheses could explain why admonishment affects false alarm rates in simultaneous lineups other than the fact that witnesses use a relative decision strategy. For example, witnesses who are admonished might false alarm less often because admonishment forces them to think more carefully about the culprit's features. That is, admonished witnesses might extract or analyze information from faces differently than those who are not admonished. As a result, they could extract information that is more diagnostic of the culprit's actual appearance. Alternatively, and more reasonably, admonishment might affect how witnesses respond to a match between what is in memory and the perceived characteristics of the face. In particular, admonished witnesses might require a more perfect match before they say one of the faces is the culprit. Believing more strongly that the culprit is in the lineup, witnesses who were not admonished might select foils that only matched the contents of their memories to a slight degree. They simply set a lower criterion that allows them to conclude that less well-matched foils are close enough in appearance to their memory of the perpetrator's looks. Still a different possibility is that admonished witnesses might be more likely to look for features that are inconsistent with their recollections of the culprit's looks. Regardless, the claim that admonishment influences people's willingness to make relative judgments requires more critical testing to determine where and how admonishment exerts its effect in the decision process. Furthermore, exactly how admonishment would discourage relative judgment is not clear

DRAFT

because the precise mechanisms involved in the relative judgment idea have received little theoretical treatment.

Self-reports of decision process

In addition to the experimental research, self-report data has been used to investigate lineup judgment strategies (Lindsay, Lea, Nosworthy et al., 1991; Lindsay & Bellinger, 1999). Subjects tend to agree that they use a relative strategy when viewing a simultaneous lineup and an absolute strategy when viewing a sequential lineup. Furthermore, these studies find that people tend to be more accurate if they report using an absolute strategy rather than a relative one. Some subjects, however, have been known to report using an absolute strategy even though the experimenter observed them comparing the lineup pictures to one another (Lindsay, 1999). Equally important, self-reports of mental process might be driven by differences between the two procedures in a way that the actual decision processes are not. For example, subjects might be more likely to report that they compare pictures in a simultaneous lineup because they can shift their visual gaze from one picture to another but can not do so in a sequential lineup. Gaze shifting might have little to do with whether subjects are using an absolute similarity standard, however. Because people respond affirmatively in laboratory experiments that they used a given mental strategy does not preclude the possibility that they compared each picture to their memories and compared the result to a criterion regardless of method of presentation.

Dual lineup procedure

Another research finding used to support the contention that relative judgment is used to select someone from simultaneous lineups is the effect that the dual lineup procedure has on accuracy rates. In the dual lineup procedure, eyewitnesses are first shown a blank lineup before they are presented with the actual lineup test. Wells (1984) found that participants who rejected the blank lineup, compared to those who picked someone out, were less likely to false alarm on a subsequently presented target absent lineup. Hit rates, however, did not differ depending on whether the participant chose someone from the blank lineup. Wells and his colleagues (1998) argued that blank lineups might be used to screen out witnesses who are prone to making relative

DRAFT

judgments. More evidence is needed, however, to demonstrate that the witnesses who false alarmed were making relative comparisons. Other than the fact that they false alarmed, we have no other evidence to indicate that those witnesses used a relative judgment process.

More importantly, the dual lineup results are also consistent with a criterion-based account. Subjects who choose someone in the blank lineup might have lower criteria (e.g., accept less of a match between their recollections of the culprit's looks and the appearance of the faces in the lineup) compared to other subjects. Thus, the dual lineup procedure might be screening out subjects who have lower criteria, not necessarily those who are inclined to make relative judgments. This competing explanation has not been empirically examined. Though not the central issue here, one also begins to wonder whether these results indicate that the tendency for relative judgment is a matter of individual differences in judgment, and not especially a problem inherent to simultaneous lineups.

Similarity

A final source of evidence taken as support for the relative judgment idea is an experimental outcome in which false alarm rates were influenced by the resemblance of the lineup members to the perpetrator (Wells, Rydell, & Seelau, 1993). In this experiment, the likelihood of participants selecting a foil from a lineup did not depend on whether only one person or all six persons in the lineup resembled the perpetrator. Presumably this is because with relative judgment someone will always be the most familiar alternative regardless of how similar the alternatives are to each other. In addition, the rate at which participants selected an innocent suspect from a target absent lineup was greater when he was the only one matching the description of the perpetrator. When only one alternative matched the perpetrator's description, he should be the most familiar and therefore the most likely alternative to be chosen. Wells et al. (1998) argued that these findings support the use of a relative judgment strategy in simultaneous lineups.

Although these results do seem consistent with the relative decision model, a criterion model can explain the same pattern of outcomes. The fact that the innocent

DRAFT

suspect was chosen more often when he was the only one matching the description of the perpetrator could be explained by the effect that the lineup similarity structure has on the odds that a given picture will exceed one's criterion. If only one person resembles the culprit, he should be more likely to surpass the criterion than any other foil. Likewise, witnesses should be less likely to pick a given foil who looks like the culprit if he is surrounded by five other persons who also look like the culprit. By chance, one would expect the more foil pictures there are in a lineup resembling the culprit, the more some witnesses would find at least one of foils to be a better match to their memories of the perpetrator than the picture designated as the innocent suspect. As a consequence, they would be more likely to pick another similar looking foil instead of the innocent suspect.

Even if one rejects this criterion explanation for the pattern of results from the Wells, Rydell & Seelau (1993) study, the fact that a recent study by Tunnicliff & Clark (2000) did not replicate this pattern raises additional concerns. Tunnicliff & Clark (2000) found no effect on hit or false alarm rates of selecting foils on the basis of similarity to the suspect or on the basis of a match to the witness's recalled description. As a result, we are left with the somewhat surprising possibility that the similarity effect reported by Wells, et al. (1993) will not replicate.

Despite some minor uncertainty, the results of empirical studies reviewed in this section seem to indicate that the rate of false alarms can be affected by a variety of experimental manipulations often with no change in hit rates. This pattern, a reduction in false alarms with little or no reduction in hit rates, has been interpreted as evidence for relative judgment processing in simultaneous lineups (Wells et al. 1998). This conclusion, however, would be justified only if an explicit and detailed model were available that described the structure and organization of the decision processes in a simultaneous lineup and other reasonable alternative models had been empirically eliminated. Unfortunately, the evidence for the use of a relative judgment process comes from a post-hoc evaluation of experiments that, in most cases, were not specifically designed to test whether a relative judgment strategy is used in simultaneous lineups. This kind of evaluation is a good place to start; but definitive policy concerning police lineup procedures ought to be based on more.

DRAFT

THE RELATIVE JUDGEMENT MODEL RECONSIDERED

Any model of how witnesses make choices from lineups will need to take account of two important empirical facts. The first is that participants in experiments reject simultaneous target absent lineups at a fairly high rate, and the second is the difference in identification performance produced by the simultaneous and sequential procedures.

The fact that witnesses frequently fail to pick anyone from simultaneous lineups suggests that they are doing more than simply picking the most familiar person from the lineup. As a result a process that allows participants to reject all members of a lineup must be an integral part of their decision strategies. Thus, a "two-process" model seems necessary to retain the idea that people use a relative decision process. However, such models can take several forms. For example, in a simultaneous lineup people might covertly select the most familiar face by comparing all familiarity values to each other. Once the most familiar face is selected, they could then compare it to an absolute decision criterion deciding whether it is familiar enough to choose. If the most familiar face exceeds the criterion, they select it. Otherwise they reject the entire lineup. Alternatively, people confronted with a simultaneous lineup could first compare each face to the same absolute criterion. If more than one face exceeds this standard, then the participants might apply a "relative" judgment or simple selection process and pick the one face that exceeds the standard by the greatest amount.[6] If no face exceeds the standard, they could reject the entire lineup. Thus, witnesses might employ a “relative” and then absolute or an absolute and then “relative” process. These different orders (and types) of strategies could produce different patterns of choice outcomes if the familiarity process tended to select a different face than the one that is most likely to be above the absolute criterion. Unfortunately, current conceptual analyses have not specified which (or whether both) of these possibilities is operating in simultaneous lineups.

One source of the lack of specificity in the relative decision strategy model has been whether "familiarity comparisons" might involve different processes and/or memory

[6] Different theorists might argue about whether the selection process among those faces that exceed the criterion is a relative one. Regardless, it clear that witnesses confronted with a lineup in which more than one item exceeds their criterion for a sufficient match would be forced to choose among them. Selecting the item with the highest similarity to the remembered culprit seems like a reasonable strategy. Whether this is called relative or absolute seems irrelevant.

DRAFT

representations than those used when deciding whether a face exceeds an absolute "identity" decision criterion. For example, familiarity decisions might be based on different memory-evidence than that used to compare each choice with whatever identity information is in memory. Specific features might be more important in judging familiarity (Burton, Bruce, & Hancock, 1999; Hancock, Burton, & Bruce, 1996) but a more "holistic" process (Cottrell, Dailey, Padgett, & Adolphs, 2001; Farah, 1996; Farah, Wilson, Drain, & Tanaka, 1998) might be involved in judging identity (or vice versa). Alternatively, features that did not match the witness's memory of the culprit might play a greater role in one procedure, whereas features that matched the contents of memory might be more important in the other procedure. Were different facial representations, e.g., surface codes and eigenvalues verses relative location of key "premorph" features (O'Toole, Wenger, & Townsend, 2001), used in familiarity judgments than in identity judgments, different specific faces might rise to the top when familiarity is the initial basis of selection than when identity is the initial basis. As a result, a final decision of the type, "Does this face look enough like the person I saw for me to pick him?" might be made to a different set of faces (across witnesses) if the faces were initially selected for familiarity than if they were selected for comparison to an identity-match criterion. While such reasoning might seem sensible, unless more detail is added (Roe, Busemeyer, & Townsend, 2001), these views cannot explain why false alarm rates and not hit rates seem to change with a change in testing procedure. After all, if familiarity-based choice processes were more likely to cause innocent look-alike examples to rise to the top when the culprit is absent, why would not the same processes cause the culprit to be more likely to rise to the top when he is present? Presumably the same familiarity-based reasons that caused an innocent look-alike to seem most familiar would also cause the culprit to seem most familiar.

Once one accepts the possibility that witnesses who are confronted with a simultaneous lineup might make decisions based on a "dual" strategy, it is only reasonable to wonder whether witnesses confronted with a sequential lineup might not use two strategies as well. For example, although the first face that witnesses see in a sequential lineup cannot be compared to faces that have yet to be seen, it is possible that the second face might be compared, in working memory, to a recalled image of the first

DRAFT

face. Similarly, the third face might be compared to recalled images of both the second and the first, and so on. If the second face is less familiar than the first one, the witness might reject it. If it is more familiar, then the witness might compare it to an absolute standard in an attempt to determine identity. Thus, relative familiarity could affect whether later faces are rejected in sequential lineups. However, the comparisons would be between images of faces in working memory with the currently presented face in the sequence.

The effect of serial position on choice rates in sequential lineups

Another issue that is made more obvious by the previous discussion is the differential importance that serial position might play in sequential as opposed to simultaneous lineups despite some reports that the positioning of the target and his replacement have no effect on choice rates in sequential lineups (Lindsay & Wells, 1985; Sporer, 1993). Although it is likely that witnesses (in the U.S.A.) will initially scan photos in a simultaneous array from left to right in a manner consistent with reading habits, they are free to look back at any photo and move their attention around in a rather "free-form" manner. However, with sequentially presented lineups, the probability that a face with a particular absolute "degree of match" to memory for the culprit will be chosen should vary with its serial position in the lineup. In particular, whether a given face that exceeds a selection standard can even be chosen depends on whether another face that exceeds the same decision standard has already been presented in the sequence. If such a face has already been presented, then the witness will not have the opportunity to choose the target's face. The later the target face appears in the sequence, the greater the number of opportunities that an earlier face will exceed the witness's "match criterion" and the less the chances that the witness will even be able to pick the target.

We conducted a Monte-Carlo simulation of a six-person sequential lineup to examine more carefully than has been done previously the theoretical implications of the effect of position on the probability that a target face will be chosen in a sequential lineup. The simulation assumed that witnesses were allowed to pick only one face and that the procedure stopped either with the first pick or after all of the faces had been seen without a pick (Lindsay, Lea, & Fulford, 1991). Thus, review of previously rejected faces

DRAFT

or multiple picks were not allowed. The simulation also assumed all pictures had a numerical value representing its recognition strength and that all five foils were selected from a unit normal distribution of "strengths" with mean zero and that the target was selected from another unit normal distribution with mean equal to d'. The target was placed in one of six different positions. Faces were presented in sequence. If a face was above a fixed criterion, that face was selected. If not, the next face in the sequence was compared to the criterion until all six faces were examined. If no face exceeded the criterion, the entire lineup ~~up~~ was rejected. The procedure was run for 4000 trials at each combination of parameter values to compute the resulting probability that a target was selected when placed in each serial position.

The results for d' values of 0 and 1 and for "memory-match" criteria placed at 0, .5, 1, and 1.5 standard deviation units (relative to the mean of the foils) are presented in Figure 2. As can be seen, the probability that the target will be selected in a sequential process depends on its serial position, d', and the location of the decision criterion. Furthermore, these results suggest that the effect of serial position increases as d' increases and decreases as the criterion becomes stricter. Thus, when the criterion was 1.5 standard deviation units above the mean of the foil distributions, position effects were minimal (the difference between choice probabilities from the first to the last position was less than .1). In fact, the position effects are so small that empirical research would require considerable power to detect such differences. The results of this Monte-Carlo simulation suggests that in sequential lineups the serial positioning of a target should have a significant effect on the probability that witnesses will select the target provided d' is not zero and/or the decision criterion very high.[7]

Insert Figure 2

Another important consequence of these simulation results is that the effect of changes in criterion on hit rate depends on the position of the target in the sequential lineup. If the target is in the 1st position, as the criterion becomes stricter, the probability

[7] Turning these results around, the lack of reported serial position effects could reflect the empirical fact that witnesses confronted with sequential lineups are setting their decision criteria very high.

DRAFT

that the target will be chosen decreases. When d' equals one, the probability of a hit decreases from .85 to .30 as the criterion increases. But when the target is in the 6th position, for the same increase in criterion, the probability of a hit increases from .03 to.26. Thus, the effect that admonishment to use a stricter criterion will have on hit rate in a sequential lineup could well depend on the target's serial position. If sequential lineups become the norm, it might be necessary to employ different criterion-setting instructions depending on the position of the target. At the very least, these simulated results make it clear that the role of serial position in sequential lineups has not yet been adequately assessed.

The results in Figure 2 were produced with the assumption that all foil distributions had a mean of zero. This is an unrealistic assumption. It seems reasonable that some foils will appear more similar, on average, to the witnesses than others. If this were the case, the effects of serial position on the expected likelihood that the target would be selected could be even more dramatic. For example, if the first foil that was seen in the sequential presentation had a d' of .5 and the target had a d' of 1, the probability that the target would be chosen when presented in the second position would be around .25 instead of the .44 shown in Figure 2. Thus, the more similar the foils are to the target in a target present sequential lineup, the more likely it is that one of the foils will be selected if presented before the target. As a result, the probability that witnesses would select the target should decrease faster with serial position.

It is also worth noting that the results in Figure 2 can represent choice rates for an innocent suspect from a target absent sequential lineup. An innocent suspect who looks more like the culprit than other foils in the lineup should be similar to a guilty target who produces lower d' values. Although the effect of serial position is smaller with lower d' values, the effect doesn't disappear even when d' is zero. As a result, if witnesses have low decision criteria, placement of the suspect in the sequence might be a critical issue in terms of controlling the rate at which witnesses will pick innocent suspects. Clearly, putting innocent suspects in later positions will decrease the odds that they will be chosen. However, putting guilty culprits in later positions will also decrease the odds that they will be chosen as well. Again, there has been virtually no discussion of the

DRAFT

potentially critical role that serial position might play were the legal system to switch to sequential lineup procedures.

Serial position in sequential lineups might produce even more complicated effects. For example, the face that is presented first may well be examined differently than faces presented second, third, and so on. The first one might set a standard against which later faces are compared. If the first face is very familiar and witnesses reject it, they might raise their standards of selection for later faces. In a simultaneous lineup, a standard is more likely to depend upon some aggregate of the entire set of faces. The effect that decision strategies such as these might have on accuracy will depend on the order of presentation of different faces and their relative familiarity and similarity to the culprit. Once again, these issues have not been discussed in the current literature.

Another potential difference between sequential and simultaneous lineups is the fact that an absolute decision criterion might change as the witness examines more faces. In particular, witnesses might set a very high standard for the first face because they want to be sure that a face that has not yet been seen is not a better example of the culprit.[8] In addition, as they progress through the faces and move to the end of the set, they might lower their standard on the grounds that they are running out of future options. The effect of moving the criterion lower as the serial position increases can be simulated in Figure 2 by moving from curve to curve as the serial position increases. The effect could be to make the serial position effect non-monotonic in comparison to holding the standard high throughout the sequence. Of course, when all of the faces are presented simultaneously, such changes in standards seem unlikely.

Given the present state of research on these issues it is not possible to decide which of several different explanations might describe the differences between simultaneous and sequential lineups. For example, although many studies verbally suggest that the position of the target did not produce significant effects, examination of the published literature yielded only two studies (Lindsay, Lea, Nosworthy et al., 1991; Sporer, 1993) that reported the actual effects of serial position on target choice rates in sequential lineups. However, in one, (Lindsay, Lea, Nosworthy et al., 1991), the target

DRAFT

was always in the same position (the last face in the sequence) and therefore serial position of the target was not varied. In addition, few studies have actually been designed to test whether the key difference between simultaneous and sequential lineups is the use by witnesses of a relative verses an absolute decision strategy or something else, e.g., where decision criteria are placed and serial position effects.

A closer look at accuracy rates

The differential effect on hit compared to false alarm rates is a key issue in deciding which of the models of the process-differences between sequential and simultaneous lineups provides the best description of witness performance. If the effect of moving from one procedure to another really has no effect on hit rates but does have an effect on false alarm rates, then not only is this important for obvious applied reasons, it also affects the details of models that might explain the differences. In particular, at first thought, a differential rate of change in false alarm compared to hit rates suggests that the difference between the two procedures cannot be due to a simple shift in a single decision criterion or even a simple difference in decision strategy. Whatever difference in strategy the change in test procedure causes for target absent lineups, one might initially believe that the same difference would apply to target present lineups, especially since witnesses do not know whether they are viewing a target present or absent lineup. For example, one could argue that the sequential procedure would induce witnesses to require greater evidence that a given face was the culprit's because witnesses might wish to withhold their choice until they saw enough faces to be sure. However, if this were the case, greater evidence would be required for target present as well as target absent lineups. Therefore one might expect a decrease in the rate of both false alarms and hits with a change from simultaneous to sequential lineup procedures. On the other hand, careful theoretical analysis might show that these expectations are not always correct.

SIGNAL DETECTION ANALYSIS OF HITS VERSUS FALSE ALARMS

Conceptual analysis of the simultaneous and sequential lineup procedures is aided by attempting to describe the differences between them in terms of signal detection theory. Figure 3 shows a signal detection representation of differences between target

[8] Were this the case, one might expect that witnesses would less confident of their rejected faces earlier

DRAFT

present and target absent lineups using distributions of "strength of stimulus-to-memory match evidence" common in signal detection theory applications to recognition memory. In this view, whether the memory is based on familiarity, feature lists, surface codes, factor structures of relationships between key features, or something else is irrelevant. Figure 3 shows an example based on a four-person (for clarity of presentation) lineup. In the target present lineup, one distribution represents the culprit and the rest represent the three remaining foils. In the target absent lineup, one distribution represents an innocent "look-alike" suspect (or some suitable replacement for the guilty culprit) and the rest the remaining foils.[9] The parameters (e.g., mean, variance, and so on) of the distributions should be the same for simultaneous and sequential lineups (unless different retrieval processes and/or memory representations are used in the two procedures) because their properties would have been determined by how well the culprit's looks were learned originally, the relative degrees to which the innocent suspect and foils looked like the culprit, and how well the culprit's picture matched the actual looks of the culprit at the time of the event.

Insert Figure 3

Before using the signal detection approach to help understand what might be happening in simultaneous and sequential lineup procedures, it is well to keep several points about this application in mind. First, unlike typical face-recognition memory experiments in which each subject sees many faces and is tested on many more, in the event-memory experiments used to assess the differences between simultaneous and sequential lineups, each participant generally sees only one target and is tested on only one lineup (either a target present or a target absent one). In addition, typically only one set of faces is used for the target present and one set for the target absent lineup. Thus, while the distributions in face-memory studies are thought to represent distributions of different face strengths "within the head" of a single subject, in event memory studies, the

than later in the sequential lineup.

DRAFT

distributions represent the strengths of the same few faces in the lineup for many different witnesses. Thus, in the event memory studies, these are individual difference distributions for a given face but in face-memory studies they are stimulus (face) distributions within each subject. Second, because the strength distributions are based on individual differences in event-memory studies, the idea that there is a single unchanging decision criterion is clearly an over simplification. If anything, a more accurate representation of the situation would assume that a distribution of decision criteria existed over the different subjects with some mean and variance.[10] Still, for ease of presentation, we can treat the mean of such a distribution of criteria as a single one that is the same for all participants.

With these issues in mind, we can use the signal detection representation in Figure 3 to help conceptually analyze what might be happening in simultaneous and sequential lineups. In particular, when confronted with a sequential lineup, witnesses might require more evidence before they would be willing to choose a particular face (in anticipation that a better face might still be available in the unseen stack) and thus their criteria would be higher on the subjective strength of evidence dimension. This assumption predicts that fewer innocent suspects (and other foils) would be chosen in a sequential than a simultaneous lineup. However, it also predicts that fewer "guilty" culprits would be chosen, as well.

On the other hand, careful analysis of the effects of moving the criterion from one location to another on expected hit and false alarm rates produces several interesting but previously ignored consequences. First, Figure 3 assumes an obscure fact about the effect of learning on recognition memory. In particular, learning not only moves the mean of a distribution of items up (on the subjective strength of evidence dimension), it also tends to increase the variance of the distribution (Ebbesen & Wixted, 1996; Ratcliff, Sheu, & Gronlund, 1992).[11] This effect on the variance of previously seen items can help explain

[9] Most, but not all, target absent lineups are constructed by replacing the target with another individual. Thus, the same foils appear in both lineups. However, other procedures have been employed in some studies.

[10] An important theoretical and empirical issue that is beyond the scope of this paper is whether the decision criterion is correlated with the strength of the target item over the subjects.

[11] One intuition that might help explain the variance increase is the idea that despite increased exposure to an item, some people might fail to encode any information about some items. As a result some people will

DRAFT

the fact that the false alarm rate decreases more than the hit rate when the decision criterion is raised. For example, in Figure 3, the change in rate of yes and no responses predicted by a shift in criterion is determined by the relative area in a distribution of items between the two criteria placements. Note how the relative area between the two criteria is larger for the innocent suspect and foil distributions than it is for the culprit distribution. As a result, an upward shift of the criterion from the simultaneous case to the sequential case would produce a greater reduction in the rate of false alarms to the innocent suspect (and to the foils) than in the rate of hits to the culprit.

Insert Figure 4

Of course, whether this effect will occur depends on the exact placements of the criteria relative to the underlying strength distributions. For example, Figure 4 shows another case with the identical strength distributions but in which both criteria are shifted to higher levels (although the distance between them on the subjective dimension is the same). As can be seen, if the criteria are placed higher on the strength dimension, it is possible that exactly the opposite result could occur. In particular, a shift from simultaneous to sequential lineups would be expected to produce a greater reduction in hits than in false alarms because there is more area under the curve between the two criteria for the former than the latter. In general, as the criteria move further upward from middle values, we would expect that equal shifts in criteria would produce increasingly larger differences in hit rates and increasing smaller differences in false alarm rates.[12]

Insert Figure 5

DRAFT

not have learned some items any better than items that people never saw before. That is, for some witnesses, a seen item might well be the equivalent to an unseen item. Other witnesses will learn those same items very well, however. Thus, the lower tail of the distribution of a previously seen item will tend to start at the same point as the distribution of unseen items but extend to values much higher than the unseen items. This will increase the variance of the strength of the seen items and well as the mean.

[12] This prediction follows directly from signal detection theory and the nature of ROC curves even in cases in which the variances of the different distributions of items are identical. The effect of increasing the variance of the studied items is to require that both criteria be higher before a shift in lineup type might cause the hit rate to change more than the false alarm rate.

These effects of criteria shifts on hit compared to false alarm rates might be more easily visualized using ROC curves. Figure 5 shows typical (e.g., equal variance of seen and not seen distributions is assumed) ROC curves for a simple "yes/no" decision task. Two different ROC curves are shown for different d' values. Superimposed on two of them are lines showing what would happen to the false alarm rate and to the hit rate if the criterion started at a lower value and then moved up such that the probability of a false alarm decreased by .1 (higher decision-criteria reduce the rate of both false alarms and hits). Of most interest is the difference in the relative size of the reduction in hit rate for a .1 reduction in the false alarm rate when the criterion is initially low (a false alarm rate of .9) and when it is initially high (a false alarm rate of .2). As can be seen, a .1 decrease in the probability of false alarms can be accompanied by a smaller change in hit rates if the criterion is initially low. However, if the criterion is initially high and then moves to a still higher value, the change in hit rate will tend to be greater than the change in false alarm rate.

Figure 6 shows predicted effects on false alarm and hit rates of shifting the criterion by varying amounts starting from either a low, middle, or high initial criterion in a simple "yes/no" decision task with d' = 1.1. As can be seen, when the criterion starts high and is shifted even higher, the signal detection model predicts a relatively greater reduction in hit than false alarm rate but when the criterion is initially low, an upward shift in the criterion tends to produce a bigger decrease in false alarm than hit rate.

Insert Figure 6

By similar reasoning, it should be clear that as the witnesses' ability to discriminate the culprit from the remaining alternatives increases (because they learn the culprit's looks better), the differential effects on hit and false alarm rates of a shift in criterion from lower to higher values will be exaggerated. Thus, if the criteria were to remain at relatively low values, then the better the witnesses learn the looks of the culprit, the bigger the decrease in false alarm compared to the decrease in hit rates. On the other hand, if the criteria were held at relatively higher values, the effect of increased learning would be to produce a bigger decrease in hit rates compared to the decrease in false alarm

DRAFT

rates. It should also be clear from the signal detection model that the mean and variance of the other distributions would affect the relative rate of hits and false alarms, as well. Simply, the closer the criteria are to the mean of the foil and innocent suspect distributions and further from the mean of the guilty culprit distribution, the more a shift will tend to affect false alarm rates compared to hit rates. The closer the criteria are to the mean of the culprit distribution and the further they are from the foil and innocent suspect distributions, the more hit rates, compared to false alarm rates, will be affected by an upward criterion shift.

This signal detection analysis of the effects of increasing the decision criterion can help explain the claim that a shift in procedure from simultaneous to sequential lineups produces a reduction in false alarm rates to target absent lineups but not in hit rates to target present lineups. If the only difference between the two procedures is that witnesses tend to set their criteria higher in sequential than simultaneous lineups, and the typical criterion placement is on the lower side in most simultaneous lineup studies, then a shift upward in the criterion would tend to produce a bigger decrease in false alarms than hits.

The signal detection analysis offers yet another factor to explain the differential effects on false alarms and hits. In particular, the presence of multiple foils adds more opportunities for false alarms than for culprit choices (assuming target present lineups contain only one target). The more foil distributions there are (i.e., the larger the lineup), the more foil items will appear to the right of the decision criterion and therefore the more total foil area there will be between the original and the new criterion (see Figures 3 and 4). In addition, the odds that at least one of the foils will have a higher strength than the target or the innocent suspect increases as the number of foils increases. These are important considerations when evaluating research that attempts to compare the results from simultaneous and sequential lineups because the majority of studies report all foil choices in target absent lineups as equivalent false alarms. In such studies innocent look-alike or suspect choices are not separately reported from non-suspect foil false alarms. As a result, a shift in criterion will necessarily produce a bigger change in the probability that any foil will be chosen in a target absent lineup than in the probability that the one culprit will be chosen in a target present lineup.

DRAFT

Monte-Carlo Simulation of a "Relative" Decision Strategy

To determine how the rate of false alarms to a target absent lineup and hits to a target present lineup would change as a function of d' and criteria placement in simultaneous lineups, we constructed a Monte-Carlo simulation of a "relative" decision strategy with a criterion. This simulation assumed that witnesses were presented with six faces. In the target absent case, all six faces were selected from a unit normal distribution with mean equal to zero. In the target present case, the target item was drawn from a unit normal distribution with mean equal to d' and the remaining foils were drawn from a unit normal distribution with mean equal to zero. The simulation first examined each alternative to determine whether it exceeded a criterion value. If none of the six items exceeded the criterion, the simulation rejected the lineup. If one or more items exceeded the criterion, the simulation examined them and chose the one that exceeded the criterion by the largest amount.[13] Once again, 4000 cases were run to determine the probabilities of all choices at d' values of .5, 1, 1.5 and 2 and for criterion values of 0, .5, 1, 1.5, 2, and 3. Figure 7 shows the results of this Monte-Carlo simulation for the probability of a false alarm (a "yes" response) for a target absent lineup compared to the probability of a hit (i.e., choosing the target) and to the probability of a "yes" response (choosing any item) for a target present lineup.

Insert Figure 7

Examination of the simulation results in Figure 7 shows that when false alarm rates (based on all choices in the target absent lineup) are compared to hit rates in a simultaneous lineup, a given increase in the criterion value will tend to produce much bigger changes in false alarm than in hit rates. For example, for a d' of 1.5, an increase in the criterion from .5 to 1.5 produces a reduction in the false alarm rate of about .55 (.89 - .34) but a reduction in the hit rate of only about .16 (.6 - .44). In other words, if researchers simply instructed witnesses to use a stricter decision criterion in a simultaneous lineup, the effect would be to produce a much larger reduction in false

[13] It is important to note that the results would be identical were the order of the two decision components reversed, i.e., pick the highest face first and then ask whether it exceeded the criterion.

DRAFT

alarm rates to target absent lineups than in hit rates to target present lineups. The results of a recent meta-analysis by Steblay (1997) of the effect of admonitions to use differentially strict decision criteria are completely consistent with this conceptual analysis. In particular, admonitions that caused witnesses to decrease their rate of false alarms in target absent simultaneous lineups had a "minimal impact" on hit rates in simultaneous target present lineups. Thus, without changing the method of presentation, instructions that caused witnesses to use stricter criteria in simultaneous lineups produced effects consistent with the signal detection analysis presented here.

On the other hand, these analyses focus on a measure of performance in target absent lineups that may lack generality, namely, a "yes" response to any foil in the target absent lineup. A more appropriate measure for generalization to the real world is the rate of innocent suspect choices in the target absent lineups. After all, the legal system generally knows that the foils cannot possibly be guilty and are only serving as distracters to test witnesses' memories (Corey, Malpass, & McQuiston, 1999; Wells & Lindsay, 1980). Unfortunately, the lack of uniformity in reporting standards and lineup design procedures means that the relevant data are simply not available for all published studies. For example, in our meta-analysis we found only 38 of the 57 lineup tests examining how adult eyewitnesses responded in simultaneous lineups reported information about suspect choices in target absent lineups.

Application of the signal detection approach to understanding the differences between the simultaneous verses sequential lineup procedures highlights a potentially important representational issue when attempting to apply theory developed in the laboratory to situations in the real world. In most, but not all, applications of signal detection to behavior, the distributions represent multiple stimuli (or trials) presented to one subject. Each subject is assumed to set a criterion for the entire set of a random mix of target present and target absent trials (unless procedures are employed to cause the subject to shift the criterion). In the "event-memory" procedures that have been used to examine the simultaneous verses sequential performance difference, each subject is shown only one "stimulus" or lineup. Typically they are not shown both target present and target absent lineups in a random mix. As a result it is possible that experimental witnesses who see a target present lineup set their criteria in different locations than

DRAFT

experimental witnesses who see a target absent lineup. Consistent with most applications of signal detection theory, the models depicted here assume that witnesses place their decision criteria in the same place for all items. However, it is conceivable that witnesses adjust their decision criteria after sampling from the items in front of them at the time that they make their decisions.[14]

Different features of simultaneous lineups might affect the criterion-setting process. For example, criteria might be adjusted based on some aggregate strength of the items. Alternatively, the item with the highest strength might affect placement of the criterion. It is also conceivable that the pattern of similarities among the faces in the lineup will affect criterion placement. Where such processes at work, it could explain how false alarm rates and not hit rates would be affected by a shift in testing procedure. For example, if witnesses tended to lower their criteria the less similar the presented faces were to their memories for the culprit's face, it would tend to cause lower criteria when the culprit was absent than when he was present in the lineup.[15] In fact, the less similar the culprit is to the innocent suspect, the more witnesses might lower their criteria when the innocent suspect and not the culprit is present.[16]

Naturally, in the sequential lineup, a different process would have to be used to set the criteria because witnesses would not have the benefit of aggregating the item similarities to which they had yet to be exposed.[17] In fact, up until the point at which the culprit or the innocent suspect was presented, the criteria would be identical in both culprit-present and innocent suspect-present sequential lineups (assuming that the foils were presented in the same order in both sequences and the key item appeared in the same position). Unfortunately, previous studies have not been designed nor the data analyzed in a manner to adequately test these alternative possibilities regarding how subjects might adjust their decision criteria across lineup type in the two test procedures.

DRAFT

[14] This view assumes that other factors, e.g., precise instructions or prior training, that might determine where witnesses place their decision criteria are vague or unclear.
[15] One reason that participants in an experiment might lower their criteria as the aggregate similarity decreased would be if they felt a "constant pressure" to pick someone regardless of the set of alternatives that they saw in front of them.
[16] The same argument would apply to the remainder of the foils in the lineup.
[17] Of course, it is possible that the criterion is set independently of the faces that witnesses see or that only the attributes of the very first face that witnesses examine has any affect on the criterion. In such cases, the

It is possible that the simple signal detection model depicted here does not describe the basic decision strategies used in these tasks. For example, guessing may be a regular part of the decision strategy. In the signal detection model, the decision is simple. Is the match between the memory representation of the face and the perception of the presented face above or below the criterion? However, witnesses might have a strategy in which they set two decision criteria. In the simultaneous lineup they might ask whether any items are above the higher of the two criteria. If only one is above this criterion, they pick it. If more than one is above the criterion, they compare their strengths and pick the highest. If none of the items are above the higher criterion, then they might determine whether any are above the next, lower, criterion. If there are some items above it, they might guess and randomly pick one of them. If no items are above this lower criterion, they simply reject the lineup.

Dual criteria strategies can be applied to sequential lineups as well. For each face, witnesses might set a high criterion above which they always pick the face and a low one below which they never pick a face. If a face falls between the two, they might guess randomly. As far as we know, more complex models such as these have not yet been applied to the simultaneous vs. sequential lineup issue because the simple relative vs. absolute distinction has been assumed to be correct.

META-ANALYSES OF SEQUENTIAL AND SIMULTANEOUS LINEUPS

It might be possible to assess some empirical consequences of the various models outlined above by examining past research outcomes. In particular, we attempted to examine the effects that the change from simultaneous to sequential testing procedures has on hit and false alarm rates by conducting a kind of meta-analysis of the results from previously published studies. Only studies that tested adult participants using simultaneous and/or sequential lineup procedures were included.

Multiple searches of the PsychInfo database (1975 to December 2000) were conducted using keywords “eyewitness,” “lineup,” and “identification.” Unpublished data provided by a recent meta-analysis conducted by Pozzulo and Lindsey (1998) were included in the analysis. Studies for which both hit and false alarm rates could not be

differences between simultaneous and sequential procedures would tend to have similar effects on criteria

DRAFT

reconstructed based on the reported data were excluded. Using these search procedures, the final sample consisted of 113 experiments from 82 papers**.** A total of 152 lineup tests were coded from this sample. Out of the 152 lineup tests, 136 tested only adult subjects and 16 tested both child and adult subjects. The characteristics of the 136 adult lineup tests were as follows: 108 were simultaneous lineups (51 included only target present conditions, 17 included only target absent conditions, and 40 included both target absent and present conditions); 11 sequentially presented lineups (4 included only target present conditions, 2 included only target absent conditions, and 5 included both target absent and present conditions); and 17 lineups were presented both simultaneously and sequentially (none included only target present conditions, 9 included only target absent conditions and 8 included both target absent and present conditions). The characteristics of the 16 lineup tests presented to both adult and child subjects were as follows: 11 were simultaneous lineups (5 included only target present conditions, none included only target absent conditions, and 6 included both target absent and present conditions); none of the experiments tested subjects using only sequentially presented lineups; and 5 lineups were presented both simultaneously and sequentially (none included only target present conditions, 1 included only target absent conditions and 4 included both target absent and present conditions). The sample captures a total of 13,198 adult eyewitness identification trials.

For obvious methodological reasons we were especially interested in comparing the hit and false alarm rates for the 12 studies that compared subject performance in simultaneous and sequential target present and target absent lineup conditions (9 studies used only adult subjects and 3 studies used both adult and child subjects). Given the relatively small number of sequential lineup studies that have been reported in the literature, we were concerned that the results might systematically vary not only because of lineup presentation mode, but also because of differences due to methodology (e.g., type of remembered event), stimuli (e.g., similarity of lineup members), and laboratory.

We examined the key issues in several different ways since the data that are available from the published reports vary in completeness. Of initial interest is whether false alarm rates in target absent lineups (e.g., a choice of any face in a target absent

placement in the target present and absent lineups.

DRAFT

lineup) and hit rates (choosing the target) in target present lineups are lower with sequential than simultaneous testing procedures. Table 1 shows the mean (raw and z-transform) proportion with 95% confidence intervals (for the former) of false alarms to target absent lineups and hits to target present lineups for all experiments in which such data for any one of the four measures were available.

Because the observations that produced these means are not strictly independent of each other, statistical analyses must be interpreted with caution. Nevertheless, a 2 x 2 unequal n's analysis of variance on the raw proportions yielded three significant effects (testing procedure $F(1,245) = 14.5$, $p = .0002$, lineup type $F(1, 245) = 6.46$, $p = .0117$, and interaction $F(1, 245) = 6.17$, $p = .0136$).[18] Examination of the confidence intervals suggest that this pattern is due primarily to the fact that the average proportion of false alarms was low in the sequential lineup procedure compared to the other three means. In fact, the residual F after removing the variance explained by a contrast comparing this mean with the remaining three was not significant.

To avoid the sampling problems inherent in the prior analyses, we examined the results for those studies in which false alarm rates for both testing procedures could be computed in the same experiment and did the same for those experiments in which hit rates could be computed for both procedures. The mean difference between the false alarm rate for the simultaneous and sequential lineups for each of the 21 published experiments was .321 (SD = .193) and was significantly different from zero ($t(20) = 7.63$, $p < .0001$). When the same computation was done for the 13 experiments that contained hit rates for target present lineups for both test procedures, the mean difference was .108 (SD = .207) and was not quite significantly different from zero ($t(12) = 1.89$, $p = .0827$) (although for a one-tailed test predicting a lower hit rate in the sequential procedure, $p = .0413$). Figure 8 shows the distributions of the actual difference scores for the experimental data. Twenty of 21 studies (95.2%) found that the false alarm rate was higher in simultaneous than sequential lineups and nine of 13 studies (69.2%) found that the hit rate was higher in simultaneous lineups.

---

[18] Results were virtually the same when z-transformed proportions were analyzed.

DRAFT

Insert Figure 8

The difference scores for the false alarm and for the hit rates in the prior analyses could not be compared directly because some hit rate differences and false alarm rate differences came from the same experiments and others did not. We found 12 experiments that reported both hit rates for target present and false alarm rates for target absent lineups for both simultaneous and sequential test procedures (Cutler & Penrod, 1988; Lindsay, Lea, & Fulford, 1991; Lindsay, Lea, Nosworthy et al., 1991; Lindsay & Wells, 1985; Lindsay et al., 1995; Lindsay et al., 1997; Melara, DeWitt-Rickards, & O'Brien, 1989; Parker & Ryan, 1993; Sporer, 1993). We used these results to compute a repeated measures analysis of variance to determine the contribution of lineup type and test procedure to the choice rates.

Insert Figure 9

Figure 9 shows the mean hit and false alarm rates (with +/- standard error of the mean bars) for the simultaneous and sequential procedure for those studies in which all four estimates were available for adult subjects. One of the first things to note about these results is that the rates of both false alarms and hits are higher for simultaneous than sequential lineups ($F(1,11) = 11.00$, $p = .007$). This is exactly what one might expect were participants using a lower decision criterion in simultaneous lineups. More importantly the significant interaction effect ($F(1,11) = 16.05$, $p = .002$) suggests that the change in testing procedure causes the false alarm rate to drop more than the hit rate, a result consistent with both the relative/absolute and signal detection/criteria shift accounts.

If we assume that whatever aspects of procedure (relative memory strengths; similarity between culprit, innocent suspect, and foils; motivation to guess; and so on) that are different from study to study are similar within each study across the two testing procedures, then the criterion shift idea makes an interesting prediction. In particular, if the only thing that is different from study to study between the simultaneous and sequential lineup procedure is the size of the shift in criterion, then as the shift in criterion

DRAFT

increases from study to study, the difference in false alarm rates between the two procedures should increase, as well. In addition, as the size of the difference in false alarm rates increases because of larger shifts in criteria placements, the size of the difference between hit rates should also increase (although the amount of increase should depend on d', average criteria placements, and serial position of the target/suspect). This idea can be deduced from examination of the effects of moving the criteria farther apart in Figure 3 or Figure 4 or by visualizing the effect on hit and false alarm rates of increasing the distance between two points on an ROC curve. Figure 10 shows the results of assessing this prediction for all of the studies in which all four estimates were available. As can be seen, the prediction appears to be supported. As the size of the difference in false alarms increased in a study, the size of the difference in hit rates also increased ($R^2 = .46$, $F(1,11) = 8.55$, $p = .02$ for the linear fit despite the fact some of the estimates were based on studies with relatively small N's and in one of the studies the faces used in the simultaneous lineup were different from those used in the sequential lineup).

Insert Figure 10

Application of signal detection theory to the differences between the two types of procedures makes another prediction, this time about changes in the rate at which witnesses choose a face (any face) from the target present compared to the target absent lineups in both procedures. In particular, when signal detection theory is applied to lineups in which a target is replaced with a look-alike suspect and the remaining foils stay the same, choice (or "yes") rates can be used to compute d' estimates (Macmillan & Creelman, 1991). The intuition behind this idea is that the presence of the target will increase the mean (and other parameters) of the presented items in the target present lineup compared to the mean of the presented items in the target absent case. The probability that the witnesses will choose someone is equivalent to the area to the right of a decision criterion under the sum of all of the face-distribution functions. Thus, the difference between the z-transformed probability of saying "yes" to the target present and the z-transformed probability of saying "yes" to the target absent lineup (regardless of

DRAFT

whether the choice was of the target, the look-alike or a foil) is the equivalent of d' in a two-alternative, yes/no, decision task. (See Figure 7 for a description of the operating characteristic results of the simulation procedure for the choice or "yes" rates in simultaneous lineups.) Equally important is the fact that the if the only difference between the simultaneous and sequential lineups is the selection procedure (as simulated in our Monte-Carlo simulations) then the two procedures will produce identical yes-response-based ROC and d' estimates. Of course, this very important prediction assumes that witnesses confronted with either lineup procedure base their choices on the same subjective evidence dimension regardless of the method used in presenting the faces. That is, it assumes that witnesses use the same evidence to judge faces regardless of procedure.

If this analysis is applied to the difference between simultaneous and sequential lineups and one assumes that the effect of moving from one to the other procedure is simply that witnesses shift their decision criteria upward, the effect will be to move the proportion of yes responses to the target present and target absent lineups in such a manner that both will move downward on linear normalized ROC curves. Figure 11 shows this prediction for different d' values. As can be seen, for different d' values, an increase in the criterion moves yes response rates for both the target present and target absent lineups down on a linear function whose perpendicular distance from the diagonal is d'. If the variances of item strengths for the target present and target absent cases are identical, then the linear functions should all have a slope of one. If the distribution for the target has a larger variance than that for the foils, then the slope should be less than one.

We tested these predictions by computing the relevant statistics for the subset of five experiments that reported enough of the relevant data to compute all four probabilities, e.g., yes to target present and target absent lineups for both the simultaneous and sequential procedures.[19] Figure 12 shows the results.

[19] It is of interest to note that the relevant data were available from published reports in only six of the set of 12 experiments in which the researchers could have reported the relevant data. In addition, we eliminated

DRAFT

Insert Figures 11 and 12

As can be seen, with the exception of the low power and somewhat peculiar post-practice results from the Parker & Ryan (1993) study in which the 12 witnesses per condition choose more often in the sequential than the simultaneous case and in which different foils were used in the target absent than the target present lineup, the results seen fairly consistent with a model that argues that a primary, if not the only, difference between the simultaneous and sequential lineups is that the witnesses tend to place their decision criteria higher when deciding in sequential compared to simultaneous lineups. Of course considerably more research is needed to verify this conclusion. Nevertheless, the pattern of results raises the likelihood that the relative v. absolute model is unnecessary to explain the differences between simultaneous and sequential lineups.

Although the "yes" response rates provide the most straightforward test of the criterion shift idea, it is possible to generalize this reasoning to hit rate for the target present lineup and total false alarm (or “yes”) rate for the target absent lineup. The operating characteristics for hit rate given target present and total false alarm rate given target absent lineups are no longer linear when normalized (Macmillan & Creelman, 1991) for the simultaneous lineup and as shall be discussed shortly depend heavily on serial position for the sequential lineup. Nevertheless, we might expect the size of the normalized differences between the hit and false alarm rates for sequential lineups to be monotonically related to the size of these normalized differences for simultaneous lineups. Figure 13 contains the results of this analysis. As the difference between the normalized hit and false alarm rates obtained for the simultaneous procedure for each experiment increased, the equivalent estimates for the sequential procedure also increased. One might not have expected this kind of regularity across experiments were witnesses using different decision strategies based on different sources of memory information about the culprit for the different presentation procedures.

one of these (Lindsay, Lea, Nosworthy et al., 1991) from the analysis because the faces used in the simultaneous procedure were different from those used in the sequential procedure.

DRAFT

Insert Figure 13

The signal detection model also makes predictions about the effect of criterion shifts on the differences between target (given target present) and innocent suspect (given target absent) choice rates. For those experiments in which the target is replaced with an innocent suspect and all of the foils remain the same, the difference between the normalized target and normalized innocent suspect choice rates for the simultaneous lineup should be monotonically related to the same difference for the sequential lineup. The exact form of the function will depend on such things as the serial position of the target/suspect in the sequential lineup and the size of the shift in criteria. Figure 14 shows the results for the six experiments in which the all of the relevant data were available. In two of the six experiments (Parker & Ryan, 1993) the foils were different in the target absent and the target present lineups and in one (Lindsay, Lea, Nosworthy et al., 1991) the faces used in the simultaneous lineup were different from those used in the sequential lineup. Nevertheless, the results in Figure 14 seem reasonably consistent with the idea that as the witnesses' abilities to discriminate a target from a suspect increased across experiments in a simultaneous presentation, their ability to discriminate the same in a sequential presentation also increased. This finding seems more consistent with the criterion shift idea than it does with a differential decision strategy idea based on different memory evidence.

Insert Figure 14

A final test of the criterion shift idea is to compare results averaged over studies directly to values that signal detection theory might predict. We attempted to perform such a test with hit and false alarm rates for the results reported in Table 1 and for the results from the 12 experiments reported in Figure 10. In each case, we compared the obtained hit and false alarm rates to the Monte-Carlo simulation results depicted in Figure 7. Recall that the latter shows how false alarm rates and hit rates would change in simultaneous lineups of size six for different d'values as the criterion shifts. Figure 15

DRAFT

reproduces these Monte-Carlo results for simultaneous lineups and superimposes on them the actual witnesses accuracy rates for simultaneous and sequential lineups. As can be seen, whether the data from all published studies (the stars) or the data from just the 12 studies in which all relevant estimates were reported (the "Xs") are used, the results seem fairly consistent with the idea that witnesses are using a stricter criteria in sequential than simultaneous lineups. The relatively larger drop in false alarm rates than in hit rates are about the same size that one might expect were witnesses simply using a stricter criterion in sequential than simultaneous lineups but everything else was the same across procedures. The fact that the sequential lineups produced hit rates very slightly higher than might be expected could be due to serial position effects, chance, the fact that some lineup studies used lineups with more than six faces, or the possibility that sequential lineups encourage witnesses to use different memory retrieval strategies.[20]

Insert Figure 15

PROCEDURAL UNCERTAINTIES IN SEQUENTIAL LINEUPS MAKE INTERPRETATION DIFFICULT

When attempting to understand the differences between sequential and simultaneous lineups it is important to consider a number of procedural issues that were left out of the guidelines (see Eyewitness evidence: A guide for law enforcement, 1999) but which may be an important part of the differences between them. Simultaneity is not the only difference in the procedures that have been employed in laboratory studies of the relative effectiveness of the two procedures (Lindsay, Lea, & Fulford, 1991).

In particular, in the sequential procedure not only are participants shown the pictures one at a time, they are also typically led to believe that they will see more than the number of potential choices shown in the comparison simultaneous lineup (generally six to eight). This is accomplished by showing the witnesses a stack of photos to look at

[20] It is also worth noting that some studies used other than six person lineups. These studies would tend to increase the size of procedure-induced shifts in false alarm rates and compared to hit rates.

DRAFT

that contains more (sometimes more than twice) that used in the simultaneous lineup.[21] Second, in most studies the witnesses seem to be told that they will see each choice only one time and that therefore they will not be able to return to a previously rejected choice.[22] Third, the actual instructions about the procedure have to be somewhat different because the procedures are different.[23] Fourth, the order of presentation of the choices and the position in which the suspect/target is placed (e.g., early in the sequence or later) might have a bigger effect on choice patterns than the position in which suspects/targets are placed in simultaneous lineups. A suspect placed in the first position in a sequential lineup might be much less likely to be picked than one placed in the sixth position because the witness might want to see other choices before making up her mind. Lastly, in sequential lineups the witnesses are generally told that they cannot pick a person and then continue viewing the remainder of the lineup (although some researchers have not followed this procedure[24], e.g., Sporer (1993). Clearly, witnesses in simultaneous lineups can tentatively choose one person and then continue to examine the remaining people.

One of the procedural uncertainties that is virtually ignored in the guide for law enforcement is a discussion of the potential effects of the serial position in the sequential lineup of the target/suspect. Possibly this is because published studies have not thoroughly examined, empirically or theoretically, the role that serial position might play in response rates. Nevertheless, as we noted earlier in this paper, serial position of the target in a sequential lineup could play a role in the likelihood that targets will be selected. To correct for this oversight, we examined the operating characteristics of simulated witnesses in sequential lineups in which the target was placed in different serial positions and then compared these results to the operating characteristics in simultaneous lineups at identical d' and criterion placements. To do this we used the data from both of the Monte-Carlo simulations described earlier.

[21] The purpose of this procedure is to ensure that a witness who has rejected the first five pictures does not then automatically assume that the last picture must be the culprit.
[22] Were they allowed to return to a previously seen item, the procedure would no longer be sequential.
[23] For example, witnesses would have to be told in the sequential procedure that they couldn't pick someone whom they have already passed. That is, after viewing face number five, they cannot then decide that the culprit was actually face number one. Such instructions are not necessary in the simultaneous lineup.

DRAFT

Figure 16 shows the results for three six-person lineups (a simultaneous lineup, a sequential lineup with the target in position 1 and a sequential lineup with the target in position 6) at three different d' values (0, 1, and 2). The figure shows the hit rate in a target present lineup plotted against the false alarm rate in a target absent lineup in which the target was replaced with a foil of equal average "memory strength" to all other foils, i.e., the suspect was not more similar to the target than any other foil. This simulation assumes that the criterion is in the same place in the target present as the target absent lineups for both procedures. It also counts a witness as making a false alarm if he/she chooses at least one of the six items presented in the target absent lineup. As a result, it follows that the total false alarm rates for the target absent lineups would be identical in the two procedures at each criterion placement. As can be seen in Figure 16, the effect on hit rate in the target present lineup as the false alarm rates increase in the target absent lineup (because the criterion is lowered) depends heavily on the serial position of the target. When the target is in the first position, the target is more likely to be chosen in the sequential procedure than the simultaneous one and the size of this difference increases as the absolute decision criterion is lowered. However, when the target is in the sixth or last position, the exact opposite result occurs. Namely, the target is less likely to be chosen in the sequential lineup than in the simultaneous lineup. Again the size of this effect increases as the criterion moves lower. In fact, at very low criteria, a target in the last position in a six person sequential lineup has a near zero probability of being chosen. The same target in the first position will be chosen nearly 100% of the time when d' is reasonably high.

DRAFT

Insert Figure 16

These results have considerable importance for interpretation of experimental studies that report total false alarm rate to the target absent lineup and hit rate to the target present lineup as their main evidence about differences between sequential and simultaneous lineups. If the present simulation-based analysis has empirical validity, it

[24] A key issue in the analysis of data from sequential lineups is how one counts correct and incorrect choices when more than one choice is made in a sequential lineup that allows witnesses to continue

will be impossible to predict the effect of testing procedure on hit rate without knowing the target's serial position, d', and criterion placement. As a result, meta-analyses of the type reported here should take account of these parameters when data from different studies are averaged. Of course, this will require that future research is designed and results are reported in a manner that provides the necessary estimates. Until they are, we may draw conclusions about differential effects of procedure on hit and false alarm rates that will not generalize to actual witnesses.

The simulation results presented in Figure 16 for hit rate given target present lineups generalize to innocent suspect choice rates given target absent lineups. If the suspect looks no more like the target than the remaining foils, then this is equivalent to a d' equal to zero. As the innocent suspect looks more like the guilty target, one might expect that the distribution of subjective memory strengths for the suspect would increase relative to the foils. As a result, the d' for the suspect should increase as the similarity between the target and the suspect increases. Thus, when compared to simultaneous suspect present lineups, suspect choice probabilities should be higher when the suspect appears in the first position in sequential lineups but lower when he appears in the last position. This effect should become more dramatic the lower the decision criteria that witnesses use.

Whether the serial position of the target or suspect is critical depends on the particular measures being used. As previously noted, several researchers (Corey et al., 1999; Wells & Lindsay, 1980) have correctly noted that false alarm rate is an inappropriate measure if the goal of eyewitness lineup research is to generalize to the real world. In the real world, a foil choice will generally be a "harmless" error from the suspect's point of view because it will not support police theories about the person they believe committed the crime. Thus, the primary focus should be on suspect choices given a target absent lineup and target choices given a target present lineup. The operating characteristics for these two measures are quite different than those presented in Figure 16. Figure 17 shows sample results from the Monte-Carlo simulations for simultaneous and sequential lineups in which we set the suspect's mean memory strength, or d', .5 above the foils and the target's mean strength, or d', 1.5 above the foils. As can be seen,

viewing the rest of the lineup after making their first choice.

DRAFT

when the target/suspect is in the very first serial position in a sequential lineup, the operating characteristics for a hit given a target present lineup and a suspect choice given a target absent lineup look very much like a typical ROC. This is not surprising. In a sequential lineup, the pictures presented after the decision to choose or not the first picture will have no effect on the odds that the witness will pick the target or the suspect. In addition, if the remaining foils are identical in both the target present and target absent lineups, they will result in equivalent rates of lineup rejections. In short, for these measures, responses to targets and suspects placed in the first serial position should be the same as those to a show-up or a two-alternative "yes/no" decision task.

The results for the simultaneous lineup are slightly different. As the criterion decreases, the rate at which target choices increase compared to the rate at which suspect choices increase slows. At the same criterion placement, both hit and suspect false alarm rates will tend to be lower for the simultaneous lineup than for the first position sequential lineups. How much lower will depend on the likelihood of at least one foil being above the criterion and having a higher strength than the target/suspect. In addition, because the probability of selecting a foil increases as the criterion decreases, a decrease in the criterion will tend to have a smaller effect on target choice rates than on suspect choice rates. The higher the criterion, the more the choice rates from simultaneous lineups should look that those from first position sequential lineups or from showups.

Finally, the results from later position sequential lineups are completely different from either first position sequential or simultaneous lineups. This is because the sequential presentation of foils before providing witnesses with the opportunity to examine the target or the suspect can greatly affect the odds that the target or suspect will even be seen. As the results in Figure 17 show, the operating characteristics for serial position 6 are quite unusual and different from those typical of ROC curves. Not only is the probability of selecting the target or the suspect considerably reduced when either is in position 6, but the effect of decreasing the criterion does not produce monotonically consistent effects on the choices rates. These rather dramatic potential effects on operating characteristics of moving from simultaneous to sequential lineups have yet to be discussed in the literature.

DRAFT

Insert Figure 17

Some researchers have argued that the most reasonable measure of lineup performance is the relative rate of choosing the guilty target from the target present lineup and the innocent suspect (who replaces the target) from the target absent lineup assuming the foils are the same in the two lineups (Navon, 1990a, 1990b; Wells & Lindsay, 1980; Wells & Luus, 1990). When this diagnosticity ratio is used rather than the simple hit and suspect false alarm rates, our simulation data, presented in Figure 18, are consistent with intuition and suggest that the effect of target/suspect position is eliminated if the target and suspect are in the same serial position in the sequential lineup. Although the absolute probabilities depend on serial position, the relative rates do not. On the other hand, in a manner somewhat inconsistent with the national guidelines, the simulation data suggest there may be very little expected difference in diagnosticity ratios between the simultaneous and sequential testing procedure if the decision criterion is at least moderately high and in the same location across procedures. However, as the simulation data presented in Figure 18 shows, at very low criterion values, simultaneous lineups would be expected to produce slightly higher diagnosticity ratios than sequential lineups when d's and criterion placement are held constant.

Unfortunately, comparisons of the diagnosticity ratios for simultaneous and sequential lineups obtained from the meta-analysis already described are somewhat inconclusive. The mean target/suspect diagnosticity for the 26 experiments for which this measure was available for simultaneous lineups was 6.1 (SD = 10.11) and for the eight experiments for sequential lineups, the mean diagnosticity was 8.0 (SD = 8.17). An independent samples t-test of the difference between these suggested that these means were far from different ($t(32) = .481$, $p = .63$). Of course such a test is not strictly appropriate because some simultaneous and sequential values came from the same experiment and some did not. When a matched-pairs t-test was computed for just those diagnosticity means that came from the six experiments containing both simultaneous and sequential lineups, the mean was again higher for the sequential lineup (10.01) than for the simultaneous lineup (3.21) but did not quite exceed traditional significance levels

DRAFT

($t(5) = 2.07$, $p = .09$). Although it appears that the diagnositicity values for sequential lineups may be higher than for equivalent simultaneous lineups, this conclusion remains somewhat tentative.

Insert Figure 18

It has been argued that diagnosticity may not be the correct measure to use to evaluate lineups (Navon, 1990a, 1990b, 1991, 1992; Wells & Lindsay, 1980; Wells & Luus, 1990) because it so obviously depends on the similarity between the innocent suspect's and guilty target's appearance and because it ignores the ecological likelihood that a suspect will look like the guilty culprit yet be innocent. Even if one rejects the criticisms of the diagnosticity measure, the simulation results presented in Figure 18 could still provide an explanation for the empirical fact that the diagnosticity ratios are higher in sequential than simultaneous lineups were future research to confirm this. In particular, the simulation data clearly show that diagnosticity increases as the criterion increases for both lineup procedures and regardless of serial position. If we are correct in claiming that witnesses tend to set their decision criteria higher in sequential than simultaneous lineups, then for this alone, diagnositicity should be higher for sequential than simultaneous lineups. Thus, it is not necessary to assume that higher diagnosticity values in sequential than simultaneous lineups means that an absolute decision strategy is being used in sequential lineups while a relative strategy is being used in simultaneous lineups. Unfortunately, an adequate test of this idea requires that criteria placement, d', target/suspect similarity, target/foil similarity, and serial position be systematically varied in an experiment that compares simultaneous and sequential procedures. Such research has yet to be done.

EXTERNAL VALIDITY ISSUES

The previous sections of this article raise several important external validity issues that we believe raise serious concerns about the willingness of some states to adopt sequential lineups and some researchers to advocate for the use of sequential over simultaneous lineup procedures at this time. In particular, it is possible that the results from laboratory simulation studies that compare error rates to simultaneous and

DRAFT

sequential lineups have ignored small but consistent effects on hit rates. That is, although not significant in most studies (possibly due to the low power of the studies), there might be a tendency for the hit rates to be lower in sequential than simultaneous lineups. Thus, it might be premature to conclude that the sequential procedure does not reduce the odds that the guilty will go free.

Second, it is possible that the beneficial effect of moving from simultaneous to sequential lineups (namely, decreasing the rate at which innocent individuals are falsely identified without decreasing the rate at which guilty culprits are identified) depends on where witnesses place their decision criteria. Based on the theoretical analyses reported here, we would only expect a greater change in innocent-suspect than guilty-target rates were there a tendency for witnesses to place their "yes/no" criteria on the lower to middle valued side of the strength dimension. Were witnesses to place their criteria high on the strength dimension for simultaneous lineups and even higher for sequential lineups, then the present theoretical analysis predicts a larger reduction in target and a smaller reduction in suspect false alarms (see Figures 5 and 17), exactly the opposite of the desired effect.

Third, selection of witnesses in the real world on the basis of their having higher confidence (than "just a guess") in the accuracy of their identifications might well be the equivalent of selecting witnesses with high criteria placements. (Ebbesen & Wixted, 1996) presented evidence that confidence estimates by eyewitnesses can be thought of as the equivalent of "yes/no" decision criteria placed at different points on the memory-strength dimension in signal detection theory. They showed that this view is consistent with results from face memory research (despite claims by experts that confidence and the accuracy of identifications are not highly correlated). If criterion placement is an important moderator of the relative difference in performance between the two lineup procedures, then before states rush to adopt sequential lineups, results from laboratory that compare performance in simultaneous and sequential lineups must present data for identification responses made with high confidence separately from those made with lower confidence. Until this type of analysis becomes routine, it is premature to assume that shift to sequential lineups will provide universal benefits with no costs. In particular, most will agree that a desirable goal of any change would be to decrease the rate at which

DRAFT

the innocent are found guilty and increase the rate at which the guilty are found guilty. In fact, the reason that the shift to sequential lineups seems so attractive is because researchers have concluded that it only affects false identifications of innocent individuals. The above analysis raises the possibility that the reported results may describe a limited case. With higher criteria, more like those that might occur in the real world, it is possible that the switch will affect hit rates more than suspect false alarm rates. Of course, different well meaning individuals might disagree about the ideal trade off between these two, e.g., how many innocent individuals is one willing to find guilty before a guilty person is let go? Regardless, the role of criteria placement in the real world compared to that in laboratory simulations is a critical issue in judging the value of the recommended switch from simultaneous to sequential lineups. Unfortunately, we currently know virtually nothing about where actual witnesses to crimes place their decision criteria.

Fourth, as our conceptual analysis of serial position effects demonstrated, the sequential presentation of alternatives may have rather unexpected effects on the operating characteristics of such lineups. As such, a switch to sequential lineups without careful instructions regarding the correct serial position in which to place the suspect might have dramatic effects on the likelihood that guilty targets will be selected (see Figure 16).

In particular, if the signal detection model is an adequate representation, then there is some probability that one or more of the foils and/or the innocent suspect will seem familiar enough to choose on some lineups. If witnesses use a relative decision strategy in the simultaneous presentation of a culprit-present lineup, then the probability that the culprit will be picked depends on the probability that the culprit is the most familiar (or has the highest subjective strength) of the alternatives. If we add the necessary absolute criterion to this process, then the probability will also depend on whether the strengths exceed the criterion as well as which has the highest strength. Regardless, it should be obvious that the culprit will be picked only if its familiarity or strength is the highest compared to the other alternatives.

DRAFT

The situation is different in a sequential lineup. In a sequential lineup, it is possible that participants might windup picking a foil, even though the culprit had the highest strength. If a foil that exceeded the absolute criterion but was lower in strength than the culprit appeared before the culprit in a sequential lineup, then the witness should pick that foil. The witness would not have the opportunity to correct this error by picking the culprit when he appeared later in the sequence because the procedure does not allow witnesses to make such corrections. This reasoning suggests that, all other things equal, the hit rates should be lower and foil choices should be higher in culprit-present sequential than simultaneous lineups. How big these differences will be should depend on how well the culprit learned (i.e., d') and how high witnesses place their absolute decision criteria.

On the other hand, there is also the possibility that one or more foils will have a greater strength than the culprit -- a probability that should increase as learning for the culprit decreases. When one or more foils have higher strength than the culprit in a simultaneous culprit-present lineup, a foil will always be chosen (assuming a simple relative decision strategy). However, in a sequential culprit-present lineup, the culprit could be chosen if it were presented before the more familiar foil. Thus, a hit might occur with the sequential procedure that would not occur with the simultaneous procedure.

We can apply the identical reasoning to innocent suspect (culprit absent) lineups. However, in this case the focus would be on innocent suspect verses foil choices. As before, whether the innocent suspect or a foil will be selected depends on several things: a) the probability of the innocent suspect having higher familiarity (strength) than one or more of the foils, b) the probability that the innocent suspect exceeds the absolute decision criterion, c) the probability that one or more of the foils exceeds an absolute decision criterion, and d) in a sequential lineup, the probability that a strong enough foil appeared before the innocent suspect. Interestingly, the first two of these would seem to depend on the similarity in appearance of the innocent suspect and the culprit. As the similarity between the suspect and culprit decreases, the likelihood that the suspect would have a higher strength than a foil should decrease.

DRAFT

It should be obvious that it will be very difficult to make predictions about the effect that different lineup procedures will have on hit and false alarm rates in the real world. Predictions would depend on the details of the nature of the lineup, whether the suspect looked a lot like the culprit, whether the witnesses learned well the culprit's looks, and where the witnesses placed their decision criteria. Although it might be argued that (Lindsay, Lea, Nosworthy et al., 1991) empirically examined some of these issues, it should be obvious that we currently do not know how variations in items strengths and strength of learning for the culprit will systematically affect hit compared to false alarm rates across the two lineup types. More importantly, we do not have standardized measures that can be applied to lineups to determine the relevant strengths of members of the lineup as well as their similarities. Finally, the research has not reported differences in hit and suspect choice rates conditional on confidence estimates or any other assessment of criterion placement.

Fifth, the signal detection analyses depicted in Figures 3 and 4 do not take into account the addition of some type of relative decision strategy that would be necessary if more than one item exceeds the criterion (no matter where that criterion is). If a relative process occurs prior to an absolute one, the number of alternatives that will be available for the latter would be unaffected by a shift in the criterion used in the latter. On the other hand, if the relative process follows the absolute one, then the number of alternatives compared in the latter will be affected by the placement of the criterion. In other words, it is possible that the difference between simultaneous and sequential lineups involves more than where decision criteria are placed. The set of alternatives from which a witness might finally choose an item could be different in the two procedures. If so, the likelihood of selecting the culprit and/or the innocent suspect would be different. Until such details are worked out, it is difficult to know how to generalize findings from the current set of studies comparing lineup type.

Sixth, it is worth noting that the application of signal detection theory to the current issue is complicated by the fact that in virtually all of the studies comparing sequential with simultaneous lineups, each subject sees only one crime and makes only one attempt to identify one culprit. As a result, the signal detection representations depicted here might be inappropriate. The testing methods do not supply each subject

DRAFT

with a large set of items. Any distribution of item strengths will completely confound subject differences with item differences because each item-strength will be in the head of a different subject. Stated differently, the distributions depicted in Figures 3 and 4 would have to represent the strength of the same item (e.g., the culprit or the innocent look-a-like) for different witnesses. As a result, it is a major simplification to assume that there is only one decision criterion for distributions of item-strengths. In fact, given the testing methods used, there are just as many criteria as item-strengths in any given study. This realization might have important applied consequences that have not yet been thoroughly analyzed.

Seventh, some might argue that the results in Figure 18, which show that the diagnosticity ratio is nearly identical for simultaneous and sequential lineups (all other things equal) and unaffected by serial position, argues that one need not be concerned about serial position effects, the lack of information about criterion placement, or the poorly defined and measured lineup similarity structure when attempting to generalize results. After all, these issues seem to go away when diagnosticity is the measure being used. On the other hand, it is important to point out that the diagnosticity measure does not take into account the odds that lineups in the real world contain culprits as opposed to innocent suspects. Thus, if most sequential lineups were to contain the target/suspect in, say, the sixth position, the diagnosticity ratio would be high (a good thing) but the likelihood that either would be selected could be quite low depending on criteria placement (see Figure 17). If most sequential lineups in the real world contained the culprit in the sixth position, then most of the time that witnesses failed to choose the suspect (because they chose a foil before getting to see the suspect), they would be failing to choose a guilty person rather than failing to choose an innocent one. Thus, whether the finding that diagnosticity seems similar across lineups is taken as evidence in support of switching to sequential lineups depends on whether one believes that most real world lineups contain innocent suspects or guilty culprits. Currently, there is very little empirical evidence on this issue, although preliminary data from our research in San Diego (Flowe, Ebbesen, Burke, & Chivabunditt, 2001) suggests that the large majority of lineups that result in prosecution contain guilty culprits.

DRAFT

Interestingly, the effect of serial position on the absolute rates of hits and suspect choices in sequential lineups would be expected to increase the lower the decision criterion and, therefore, the higher the rate of false alarms. Thus, as the results in Figure 16 suggest, serial position effects are greatest when false alarm rates are very high. Thus, if one believes, contrary our (Flowe et al., 2001) preliminary data that false alarms occur at a high rate in the real world, then the importance of serial position effects is enhanced. If, however, one believes that false alarms are relatively rare events in the real world because witnesses tend to set their criteria relatively high, then serial position effects are less worrisome and might even be irrelevant.

Eighth, potential problems that may arise in the field once police officers actually perform the sequential procedure have not been given adequate consideration. For example, what would happen if a witness was to say when shown number 2, that she thinks it is the suspect, but would like see the rest of the pictures to be sure? Will an officer in the field allow her to see the rest of the pictures and then come back and pick #2? Furthermore, in the field it is plausible that real witnesses will want to see some or all of the pictures again. There is evidence suggesting, however, that allowing subjects to view a sequentially presented lineup twice increases the rate of false alarms (Lindsay, Lea & Fulford, 1991) and some (Parker & Ryan, 1993) that suggests practice reduces false alarms. How will police under pressure to solve a criminal investigation react to such requests?

Lindsey and Bellinger (1999) recently investigated the effectiveness of lineup presentation strategies used by the police in Ontario, Canada. Police in this district were allowing eyewitnesses to self-administer sequential lineups to make sure that identification outcomes would not be biased by the officer conducting the identification. Apparently police were attempting to circumvent a guideline for administering lineups indicating that the officer giving the lineup to the witness should be blind to the identity of the suspect. Subjects in this study who self-administered their own lineups were frequently observed violating the experimenter's instructions to not re-examine or compare photographs. Furthermore, all of the subjects who compared photographs failed to reject the target absent lineups, even if they reported using an absolute rather than a relative judgment strategy. Though its not clear whether real eyewitnesses would also fail

DRAFT

to follow such instructions, which were designed to discourage relative judgment, the results of this study underscore some of the difficulty and dangers that the police are likely to encounter when attempting to follow the guidelines suggested by the NIJ workgroup.

KEY LIMITATIONS OF THE CRITERION MODEL

The signal detection based analyses reported here suffer from two potential limitations that require further exploration in future research. One concerns the already mentioned issue that the assumed distributions of memory strengths are not item distributions in the head of single witnesses but item distributions across different witnesses. As such, assumptions about the form of these distributions based on past research with signal detection might be incorrect. For example, the Monte-Carlo simulations assumed normal distributions with skew and kurtosis equal to zero. Actual witness distributions might be quite different. However, many of the predictions made by the Monte-Carlo simulations depend on the fact that the distributions of strengths are normal. To the extent that they are not, the predictions will be different. On the other hand, virtually nothing is known about such distributions, even from laboratory research, because information about them has not been presented in published studies.

The second problem concerns the fact that criterion placements will be different for different witnesses and although there is some reason to believe that individual differences in criterion placement and face memory are not highly correlated (Ebbesen & Wixted, 1996), the nature of criteria placement across actual witnesses is clearly something about which we currently know very little.

SUMMARY

The results of the meta-analyses and simulations presented here are consistent with the idea that a major difference between the sequential and simultaneous lineup procedure is that witnesses set a higher criterion for a match between their recollection of the culprit and the pictures in the sequential than the simultaneous lineup. We are not suggesting that this the only decision-process difference between the two procedures. We are suggesting, however, that the psychological processes involved in eyewitness identification with these procedures require further study, both conceptually and

DRAFT

empirically, before the sequential lineup is uniformly recommended as the preferred identification procedure. Independent of whether one accepts the criterion-shift model as correct, the present analyses raise the theoretical importance of empirical predictions that have been ignored up to now. More empirical work is clearly needed to establish the judgment processes that eyewitnesses might use in evaluating sequential and simultaneous lineups. Furthermore, the effect of using a sequential over a simultaneous procedure on accuracy rates needs to be examined under a variety of witnessing and instructional conditions. The criterion shift decision model predicts that whether a sequential or simultaneous lineup should be used depends on multiple factors that were virtually ignored in the NIJ published guidelines and supporting publications (e.g., (Wells et al., 1998)). Even if this model is eventually proven wrong, its plausibility and its specificity should raise concern among those who want to replace simultaneous lineup procedures with sequential ones.

DRAFT

REFERENCES

Behrman, B. W., & Davey, S. L. (1999). Eyewitness memory for actual crimes: An archival analysis. Paper presented at the American Psychological Society, 11th Conference, Denver, CO.

Behrman, B. W., & Vayder, L. T. (1994). The biasing influence of a police showup: Does the observation of a single suspect taint later identification? Perceptual & Motor Skills, 79(3, Pt 1), 1239-1248.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. Cognitive Science, 23(1), 1-31.

Corey, D., Malpass, R. S., & McQuiston, D. E. (1999). Parallelism in eyewitness and mock witness identification. Applied Cognitive Psychology, 13(Spec Issue), S41-S58.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? The view from UCSD. In J. T. Townsend (Ed.), Computational, geometric, and process perspectives on facial cognition (pp. 347-395). Mahwah, NJ: Lawrence Erlbaum Associates.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. Journal of Applied Psychology, 73(2), 281-290.

Ebbesen, E. B., & Wixted, J. (1996). A signal detection analysis of the relationship between confidence and accuracy in eyewitness memory. La Jolla: University of California, San Diego.

Eyewitness evidence: A guide for law enforcement. (Research report)(1999). Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Farah, M. J. (1996). Is face recognition "special"? Evidence from neuropsychology. Behavioural Brain Research, 76(1-2), 181-189.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? Psychological Review, 105(3), 482-498.

Flowe, H., Ebbesen, E. B., Burke, C., & Chivabunditt, P. (2001). At the scene of the crime: An examination of the external validity of published studies on line-up identification accuracy. Toronto, Canada: American Psychology Society Annual Convention.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. Journal of Personality & Social Psychology, 64(4), 525-537.

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. Memory & Cognition, 24(1), 26-40.

Levi, A. M. (1998). Protecting innocent defendants, nailing the guilty: A modified sequential lineup. Applied Cognitive Psychology, 12(3), 265-275.

Lindsay, R. C., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. Journal of Applied Psychology, 76(5), 741-745.

Lindsay, R. C., Lea, J. A., Nosworthy, G. J., & Fulford, J. A. (1991). Biased lineups: Sequential presentation reduces the problem. Journal of Applied Psychology, 76(6), 796-802.

DRAFT

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. Journal of Applied Psychology, 70(3), 556-564.

Lindsay, R. C. L. (1999). Applying applied research: Selling the sequential line-up. Applied Cognitive Psychology, 13(3), 219-225.

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. Journal of Applied Psychology, 84(3).

Lindsay, R. C. L., Craig, W., Lee, K., Pozzulo, J. D., Rombough, V., & Smyth, L. (1995, July). Eyewitness identification procedures for use with children. Paper presented at the Society for Applied Research in Memory and Cognition, Vancouver, British Columbia.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., & Lee, K. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. Law & Human Behavior, 21(4), 391-404.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York, NY, USA: Cambridge University Press.

Melara, R. D., DeWitt-Rickards, T. S., & O'Brien, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. Journal of Applied Psychology, 74(5), 706-713.

Navon, D. (1990a). Ecological parameters € nonlineup evidence: A reply to Wells and Luus. Journal of Applied Psychology, 75(5), 517-520.

Navon, D. (1990b). How critical is the accuracy of an eyewitness's memory? Another look at the issue of lineup diagnosticity. Journal of Applied Psychology, 75(5), 506-510.

Navon, D. (1991). "Ecological parameters € nonlineup evidence: A reply to Wells and Luus": Correction. Journal of Applied Psychology, 76(3), 407.

Navon, D. (1992). Selection of lineup foils by similarity to the suspect is likely to misfire. Law & Human Behavior, 16(5), 575-593.

O'Toole, A. J., Wenger, M. J., & Townsend, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In E. James T. Townsend (Ed.), Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. (pp. 1-38): Mahwah, NJ, US.

Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. Law & Human Behavior, 17(1), 11-26.

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. Journal of Applied Psychology, 84(2).

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. Psychological Review, 99, 518-535.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. Psychological Review, 108(2).

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78(1), 22-33.

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In J. D. R. M. P. T. E. David Frank Ross (Ed.),

DRAFT

Adult eyewitness testimony: Current trends and developments. (pp. 300-327): Cambridge University Press, New York, NY, US.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. Law & Human Behavior, 21(3), 283-297.

Tollerstrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In J. D. R. M. P. T. David Frank Ross (Ed.), Adult eyewitness testimony: Current trends and developments. (pp. 144-160): Cambridge University Press, New York, NY, US.

Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? Law & Human Behavior, 24(2), 231-258.

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Social Psychology, 14(2), 89-103.

Wells, G. L. (1993). What do we know about eyewitness identification? American Psychologist, 48(5), 553-571.

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. Psychological Bulletin, 88(3), 776-784.

Wells, G. L., & Luus, C. E. (1990). The diagnosticity of a lineup should not be confused with the diagnostic value of nonlineup evidence. Journal of Applied Psychology, 75(5), 511-516.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. Journal of Applied Psychology, 78(5), 835-844.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. Law & Human Behavior, 22(6).

Yarmey, A. D. (1998). Person identification in showups and lineups. In et al. (Ed.), Eyewitness memory: Theoretical and applied perspectives. (pp. 131-154): Mahwah, NJ, USA.

Yarmey, A. D., & Morris, S. (1998). The effects of discussion on eyewitness memory. Journal of Applied Social Psychology, 28(17), 1637-1648.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identification in showups and lineups. Law & Human Behavior, 20(4), 459-477.

DRAFT

Table 1

Mean proportion (over all experiments in the sample) of false alarms to target absent lineups and hits to target present lineups for sequential and simultaneous lineup procedures from which the proportions could be computed.

Procedure	Lineup	N[a]	Mean	95% Confidence	Std Error	Mean z[d]
Sequential	Target Absent	28	.293[b]	.224-.361	0.033	.265
Simultaneous	Target Absent	84	.486[b]	.442-.530	0.022	.496
Sequential	Target Present	23	.446[c]	.363-.529	0.042	.431
Simultaneous	Target Present	114	.488[c]	.450-.525	0.019	.482

[a] Number of experiments contributing to proportion (not number of subjects).
[b] Proportion of total false alarms
[c] Proportion of hits
[d] Mean proportions based on average of z transformations of proportions.

DRAFT

Figure Captions

Figure 1. Choice rates from Wells (1993) study of the effects of removing and not replacing a target in simultaneous lineups.

Figure 2. Predicted effects of d', decision criterion placement, and serial position on the probability [p(h|tp)] that a target will be selected from a sequentially presented six item target present lineup. The results are based on a Monte Carlo simulation of an absolute decision strategy in which all items are selected from a unit normal distribution. All foil distributions have a mean of zero and the target has a mean of d'. The simulation assumes that criterion placement is unaffected by serial position.

Figure 3. Signal detection representation of foil, target, and innocent suspect "strength of memory" distributions in four-person target present and target absent lineups. The representation shows the effect of moving a decision criterion from a lower value in simultaneous lineups to a higher one in sequentially presented lineups.

Figure 4. Signal detection representation of foil, target, and innocent suspect "strength of memory" distributions in four-person target present and target absent lineups. The representation shows the effect of moving a decision criterion from a lower value in simultaneous lineups to a higher one in sequentially presented lineups when the criteria are placed relatively high compared to representation in Figure 3.

Figure 5. Signal detection representation of the predicted effect of a shift in criterion placement on the probability of a hit given a seen item and on the probability of a false alarm given a not seen item in a "yes/no" choice procedure as a function of d' and whether the criteria start initially low (e.g., so the false alarm rates are high at around .9) or initially high (so the false alarm rates are around low at around .2). When the criteria are initially low, a shift in criteria will tend to produce a larger change in false alarm than hit rates but when the criteria are high, the opposite is true. The larger d' the bigger the differential effect is.

Figure 6. Signal detection predicted decrease in false alarm and hit rates in a simple "yes/no" recognition memory task with d' = 1.1 as a result of shifting the decision criterion a fixed amount. The shift either begins with the criterion set to a low value, a

DRAFT

middle value, or a high value. When the criterion is initially low, a given shift upward in the criterion will tend to produce a bigger decrease in false alarms than hits. But when the criterion is initially high, a shift upward in the criterion will tend to produce a bigger decrease in hits than in false alarms.

Figure 7. Results from a Monte-Carlo simulation of choices made to a simultaneous lineup in which witnesses decided whether a face was above a criterion. If more than one was above the criterion, the one that exceeded the criterion by the largest amount was selected. Solid data points represent results for the probability of a hit given a target present lineup (h|tp) and probability of a false alarm given a target absent lineup (fa|ta) at each criterion value. The open data points represent results for the probability of a "yes" response to the target present lineup (yes|tp) and the probability of a false alarm (also any "yes" response) given a target absent lineup at the same criterion values. Note how the effect of variation in criterion placement (movement along a given curve) is much bigger on false alarm rate (measured as a choice of any foil in the target absent lineup) than on hit rate over a wide range of criterion values.

Figure 8. Differences in hit (Xs) and false alarm rates (open circles) produced by changing from a simultaneous to sequential method of presentation for all experiments in which hit or false alarm data were available for the two methods. Jiggle in the x-axis was added to make it easier to see overlapping data points.

Figure 9. Mean proportion of false alarms to target absent lineups and hits to target present lineups in the 12 experiments in which both of these were available for simultaneous and sequential lineups. Standard error of the mean bars are shown. Both hit and false alarm rates tended to be lower in the sequential than the simultaneous procedure, however, this effect was bigger for the false alarms than hits.

Figure 10. Relationship between the size of the simultaneous minus sequential difference in hit rate and in false alarm rate across the twelve different experiments in which the relevant data were available. As the size of the difference in hit rate increased, the size of the difference in false alarm rate also increased as would be expected were the differences produced by a shift in the decision criterion.

DRAFT

Figure 11. Predicted effect of criterion shift on normalized probability of saying "yes" to target present and target absent lineups for both simultaneous and sequential lineups.

Figure 12: Relationship between z-score probability of witnesses saying “yes” to a target absent lineup up and “yes” to a target present lineup in simultaneous and sequential lineup testing procedures for the five experiments in which the relevant data were available (1 = Parker & Ryan, 1993, after practice; 2 = Melara, et al., 1989; 3 = Lindsay, Lea, & Fulford, 1991, Exp. 1; 4 = Lindsay & Wells, 1985; 5 = Parker & Ryan, 1993, without practice).

Figure 13. Relationship between the differences in the z transformed hit rate to target present lineups z(H|TP) and the z transformed false alarm rate z(FA|TA) for simultaneous and sequential procedures. Filled circles represent the data for those experiments in which the target was removed and replaced with a suspect for the target absent lineups. Xs represent those studies in which some other procedure was used to construct target absent lineups. As the difference between hit and false alarm rates increased for simultaneous lineups, the same difference increased for sequential lineups across the available experiments.

Figure 14. Relationship between the differences in the z transformed hit rate to target present lineups [z(target|tp)] and the z transformed suspect-only false alarm rate [z(sus|ta)] for simultaneous and sequential procedures. Circles represent the data for those experiments in which the target was removed and replaced with a suspect for the target absent lineups. The best fitting linear function is shown for these data points. Xs represent those studies in which some other procedure was used to construct target absent lineups. As the difference between target and suspect choice rates increased for simultaneous lineups, the same difference increased for sequential lineups across the available experiments.

Figure 15. Comparison of results from Monte Carlo simulation and from experiments. Stars represent results from all of the available experiments (see Table 1). Xs represent data from the 12 experiments in which all four estimates were available (see Figure 9). The two data points with the higher false alarm and higher hit rates report the average results from simultaneous lineups and the other two are for sequential lineups. Grayed

DRAFT

functions are the results from the Monte Carlo simulation of simultaneous lineup procedures reported in Figure 7.

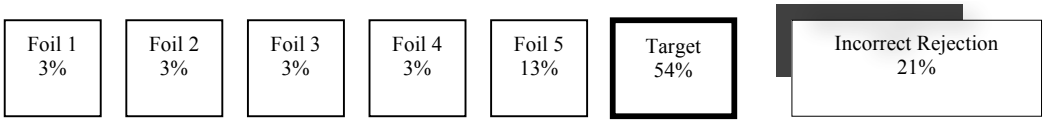
Figure 16. Operating characteristics (hit rate given a target present lineup and false alarm rate given a target absent lineup) based on Monte Carlo simulations as a function of d', lineup procedure, and serial position (1^{st} and 6^{th}) in the sequential case. In the simulations all foil distributions that appeared in the target present lineup had mean evidence strengths of zero. The target was replaced with a foil with mean evidence strength also equal to zero. Position refers to the serial position of the target in the target present sequential lineup.

Figure 17. Operating characteristics (derived from Monte Carlo simulations) for hit and suspect false alarm rates in six-person simultaneous and sequential lineups with the target/suspect either in serial position 1 or serial position 6. The d' for the target was set at 1.5 standard deviation units above the target present foils and the d' for the suspect was set at .5 standard deviation units above the target absent foils. The foils in the target absent and target present lineups were all set to a mean of zero and variance equal to 1.

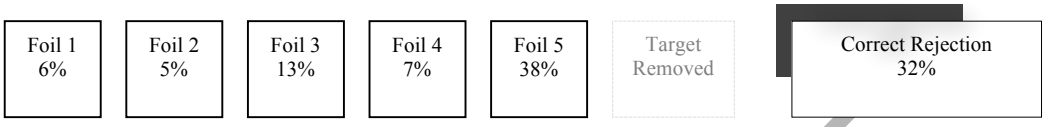
Figure 18. Effect of d' (mean of target distribution compared to zero mean of all foil distributions) and criterion placement on Monte Carlo simulation-based diagnosticity estimates (probability of target choice given target present lineup divided by probability of suspect choice given target absent lineup) for simultaneous and sequential lineups. Suspect d' was set at .5 for the results in this figure. Position of the target/suspect in the lineup has no effect the diagnosticity of the sequential lineup. Simultaneous and sequential lineups produce nearly identical diagnosticity results when the decision criteria are identical and reasonably high.

DRAFT

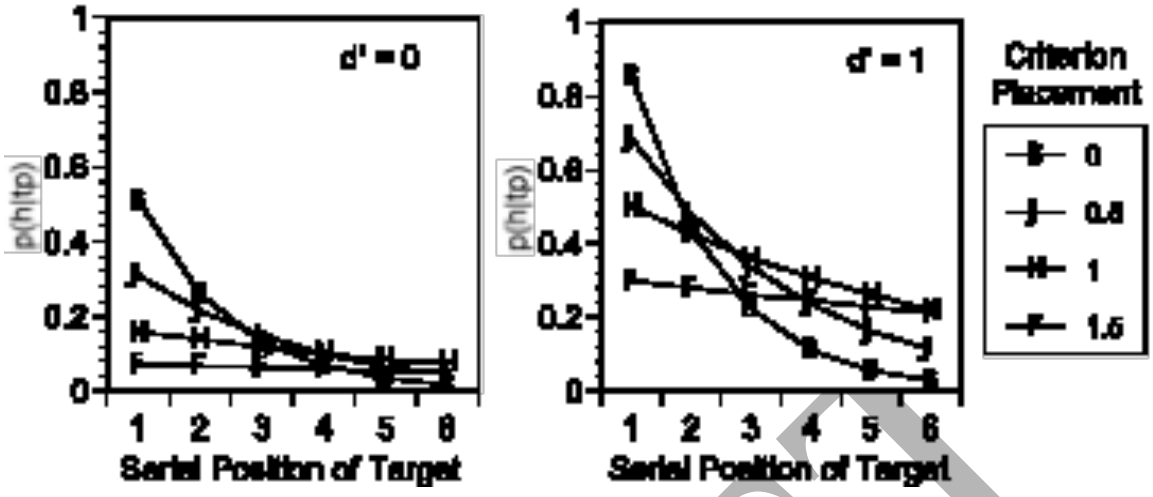
Group 1: Distribution of Choices in Target Present Lineup.



Group 2: Distribution of Choices in Target Absent (Target Removed and Not Replaced) Lineup.

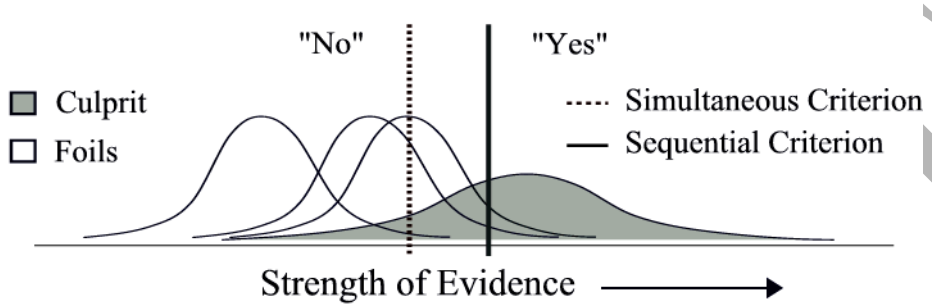


DRAFT

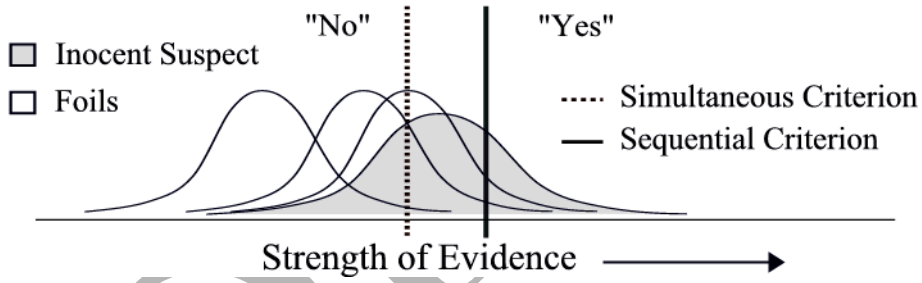

d' = 0
d' = 1
p(h|tp)
Serial Position of Target
Criterion Placement
0
0.5
1
1.5

DRAFT

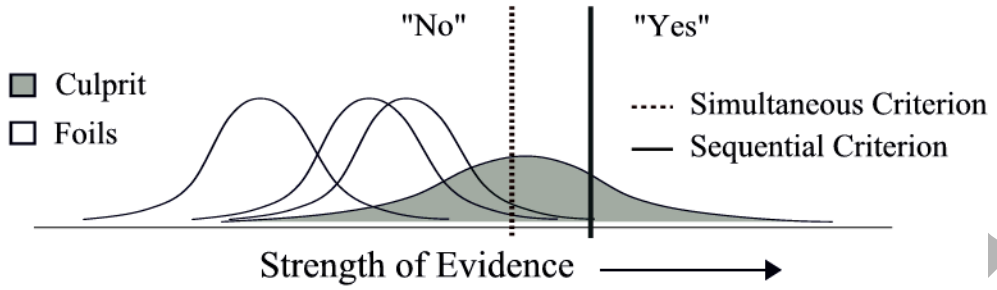
Target Present Lineups



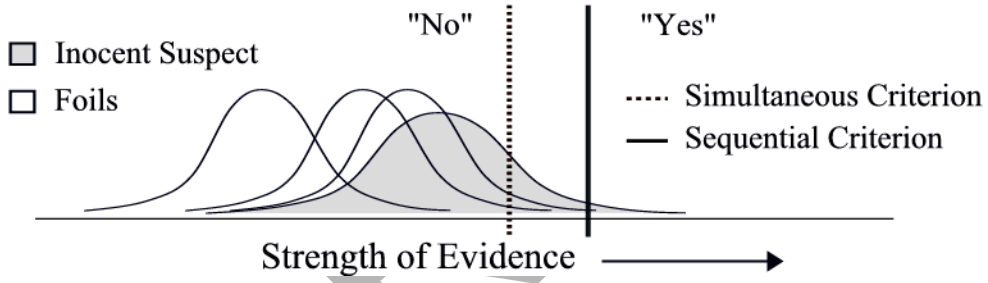
Target Absent Lineups



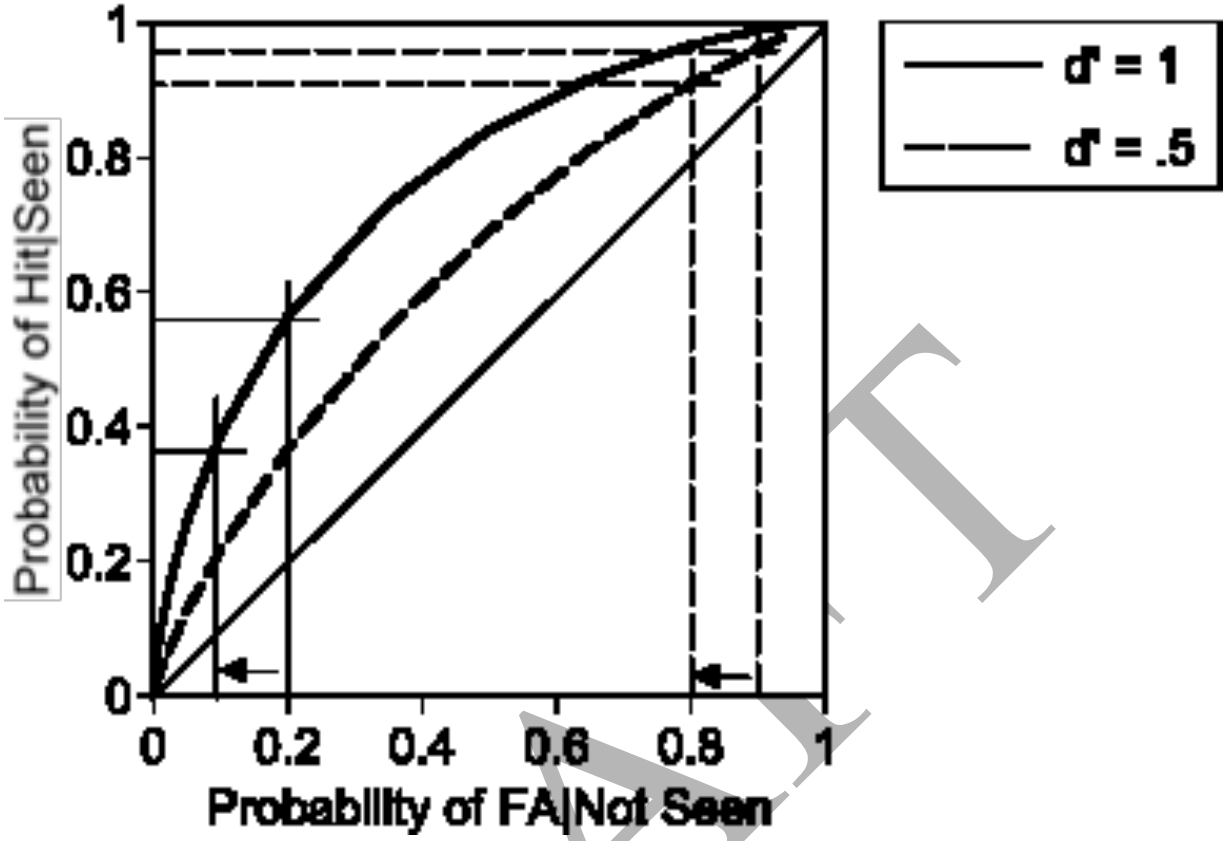
Target Present Lineups



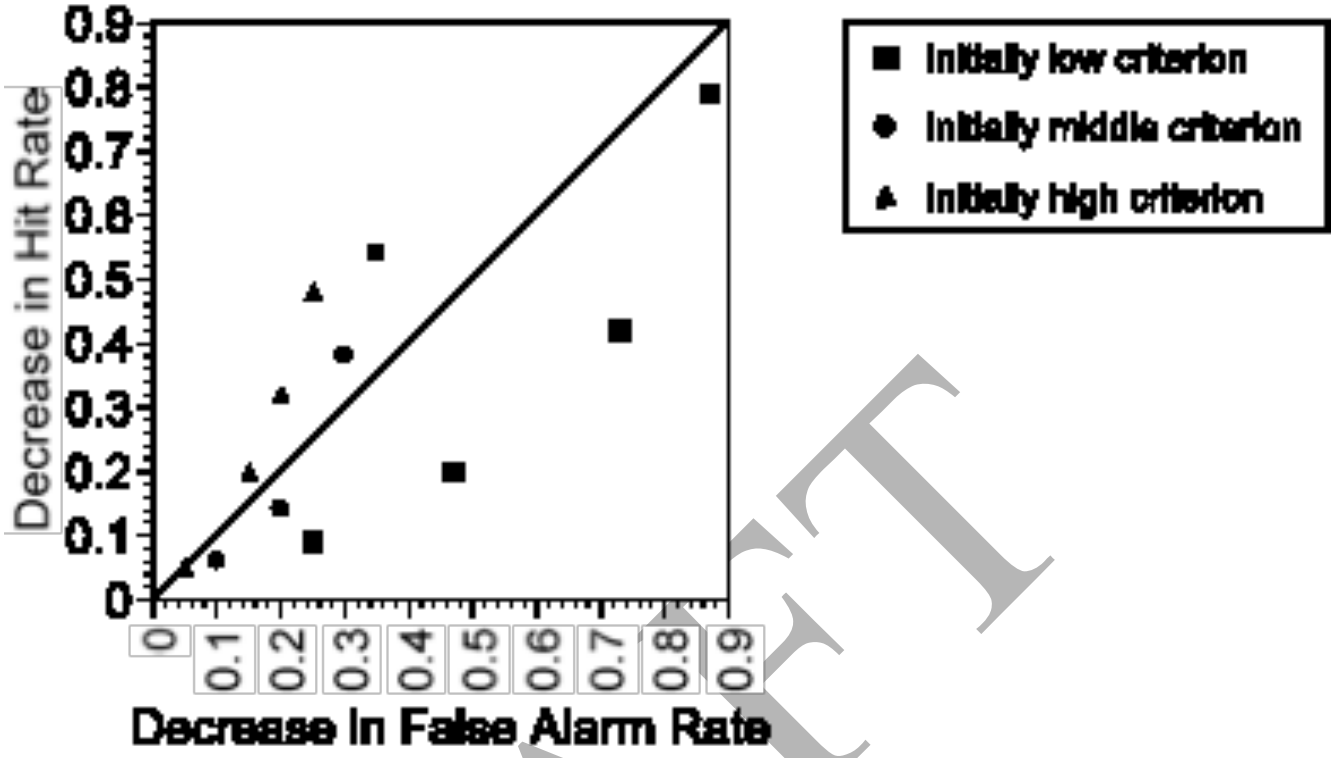
Target Absent Lineups



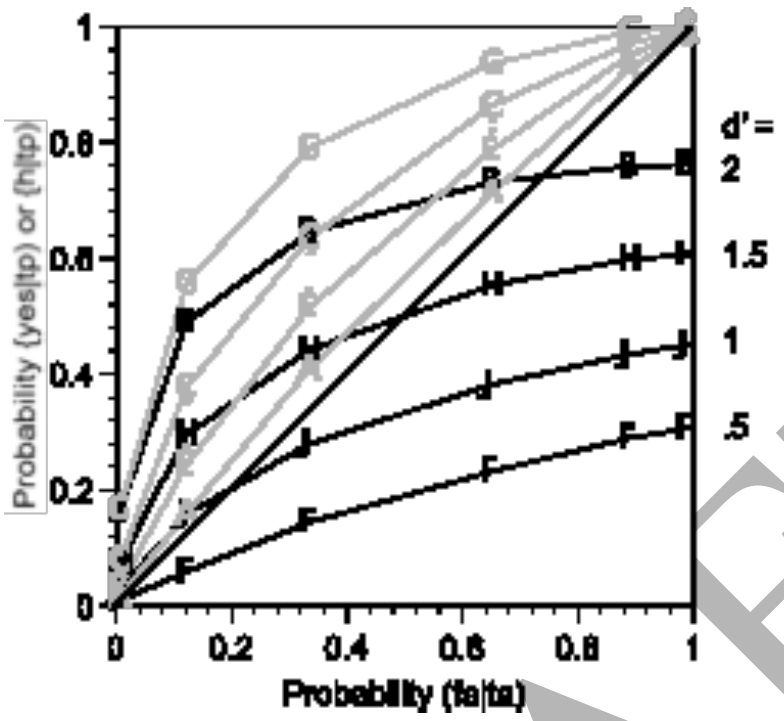
DRAFT


d' = 1
d' = .5
Probability of Hit|Seen
Probability of FA|Not Seen
0
0.2
0.4
0.6
0.8
1

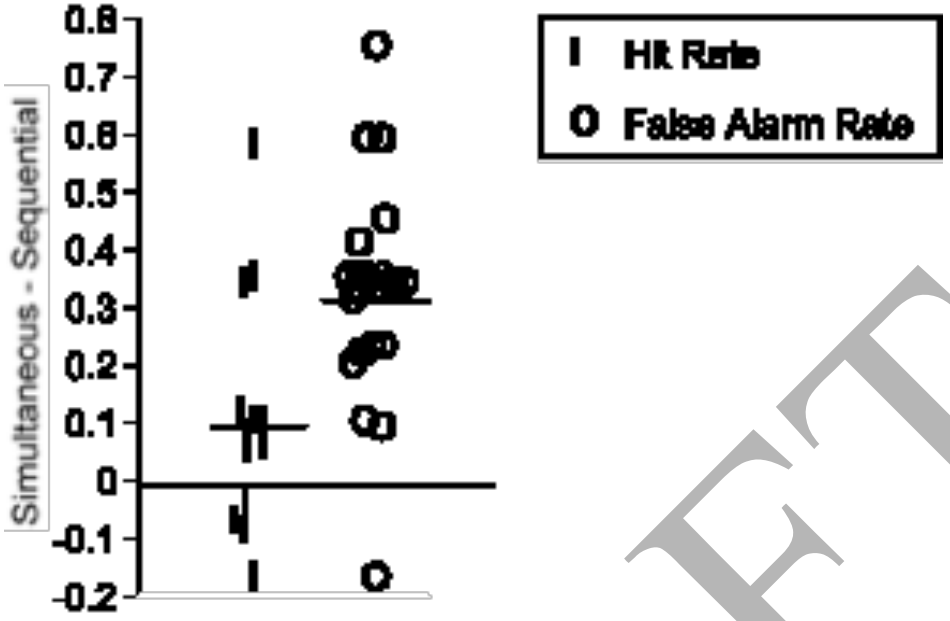
DRAFT


Decrease in Hit Rate
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Decrease in False Alarm Rate
Initially low criterion
Initially middle criterion
Initially high criterion

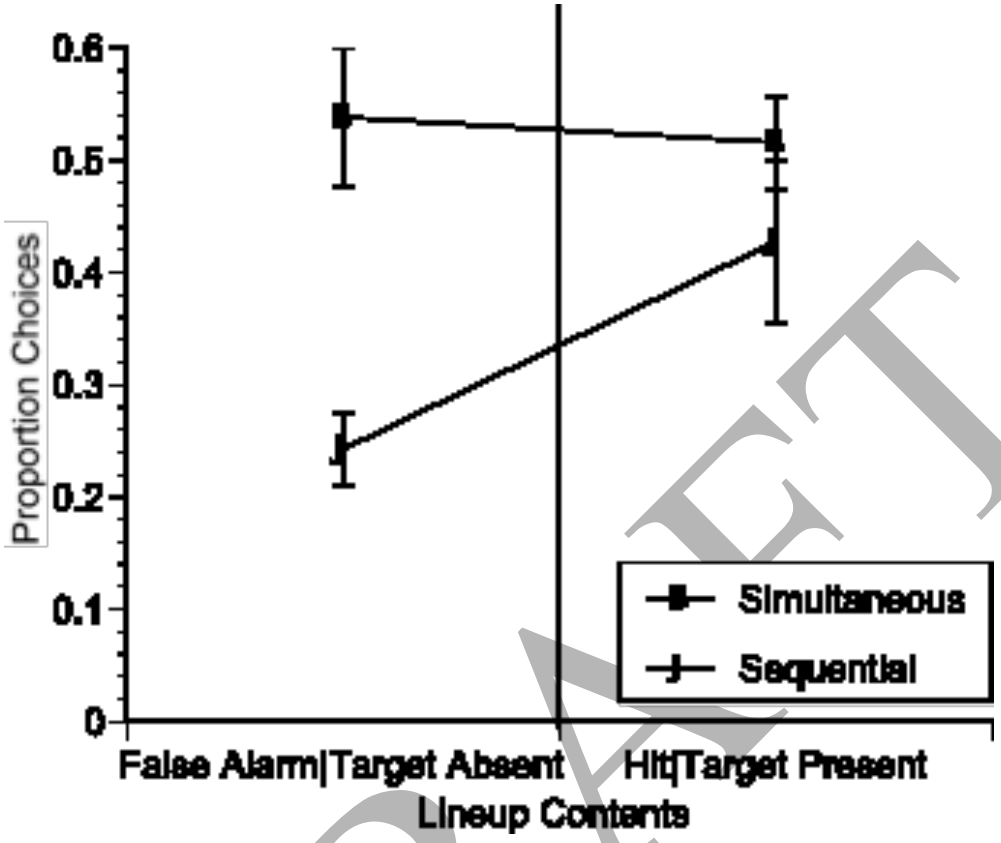
DRAFT


Probability (yes|tp) or (h|tp)
Probability (fa|ta)
0
0.2
0.4
0.6
0.8
1
d' =
2
1.5
1
.5

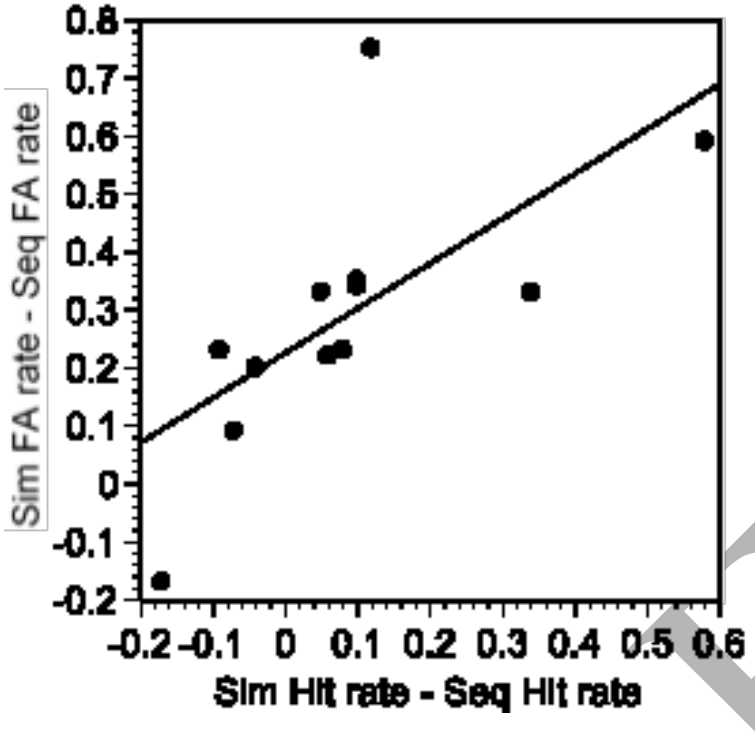
DRAFT


Simultaneous - Sequential
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
Hit Rate
False Alarm Rate

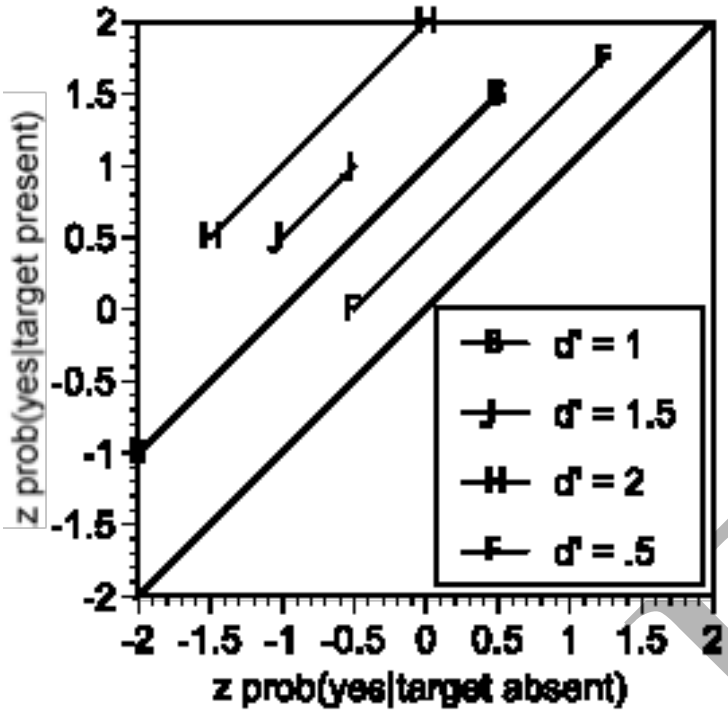
DRAFT


0.6
0.5
0.4
0.3
0.2
0.1
0
Proportion Choices
Simultaneous
Sequential
False Alarm|Target Absent
Hit|Target Present
Lineup Contents

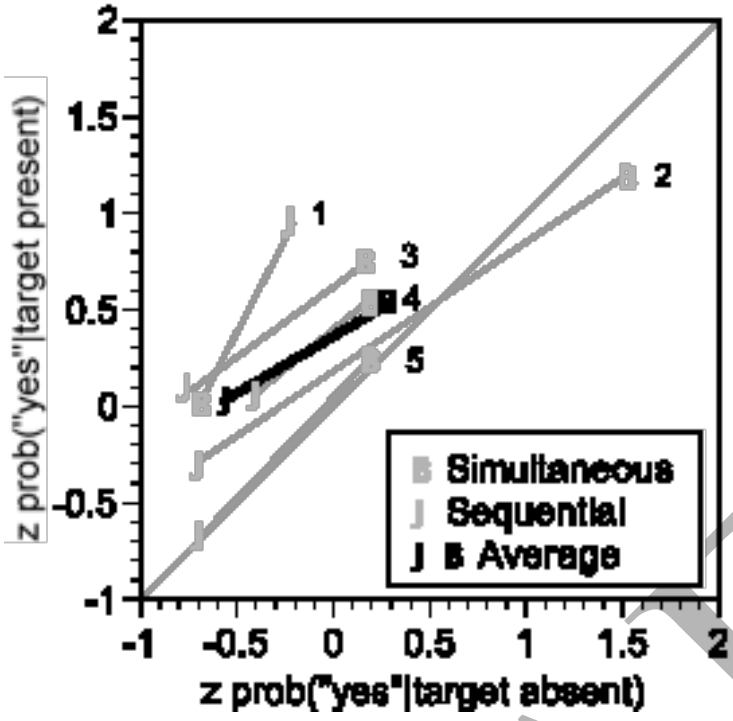
DRAFT


Sim FA rate - Seq FA rate
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6
Sim Hit rate - Seq Hit rate

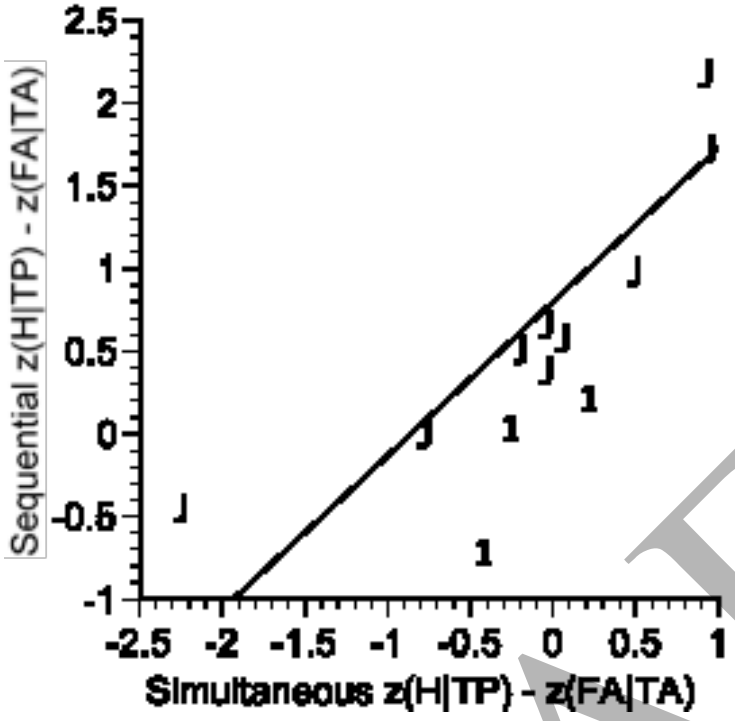
DRAFT


z prob(yes|target present)
z prob(yes|target absent)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2 -1.5 -1 -0.5 0 0.5 1 1.5 2
H
F
B
J
H
J
F
B
d' = 1
d' = 1.5
d' = 2
d' = .5

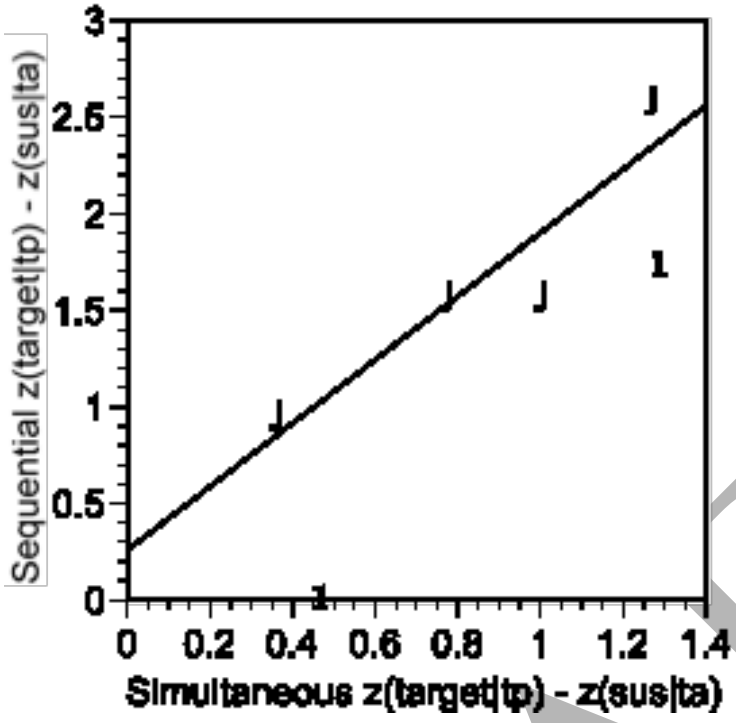
DRAFT


z prob("yes"|target present)
z prob("yes"|target absent)
2
1.5
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1
1.5
2
1
2
3
4
5
Simultaneous
Sequential
Average

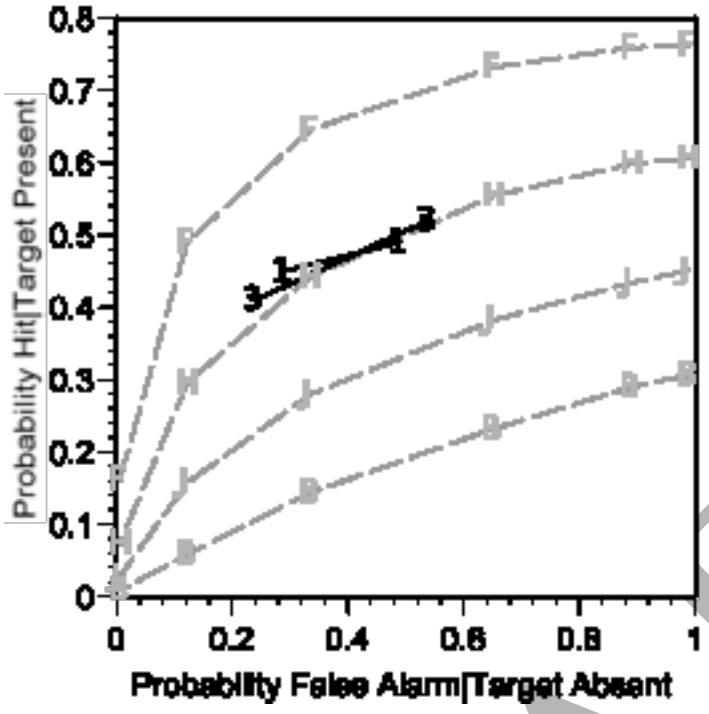
DRAFT


Sequential z(H|TP) - z(FA|TA)
Simultaneous z(H|TP) - z(FA|TA)
2.5
2
1.5
1
0.5
0
-0.5
-1
-2.5
-2
-1.5
-1
-0.5
0
0.5
1

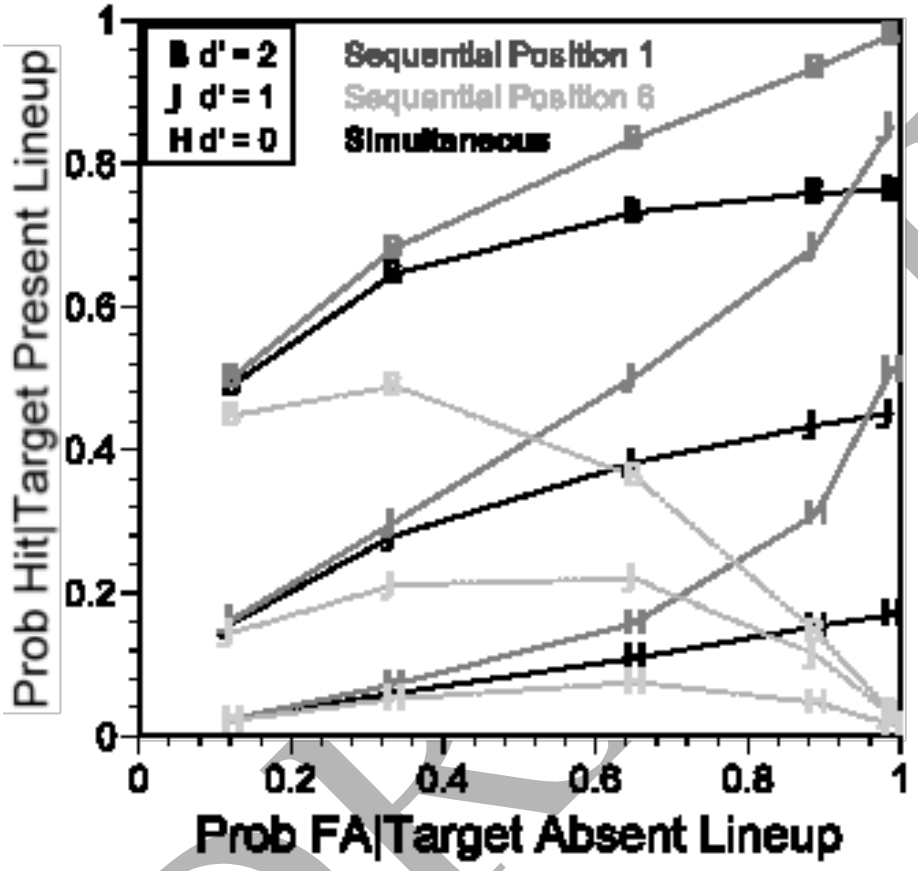
DRAFT


Sequential z(target|tp) - z(sus|ta)
Simultaneous z(target|tp) - z(sus|ta)
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
J
J
J
1
1

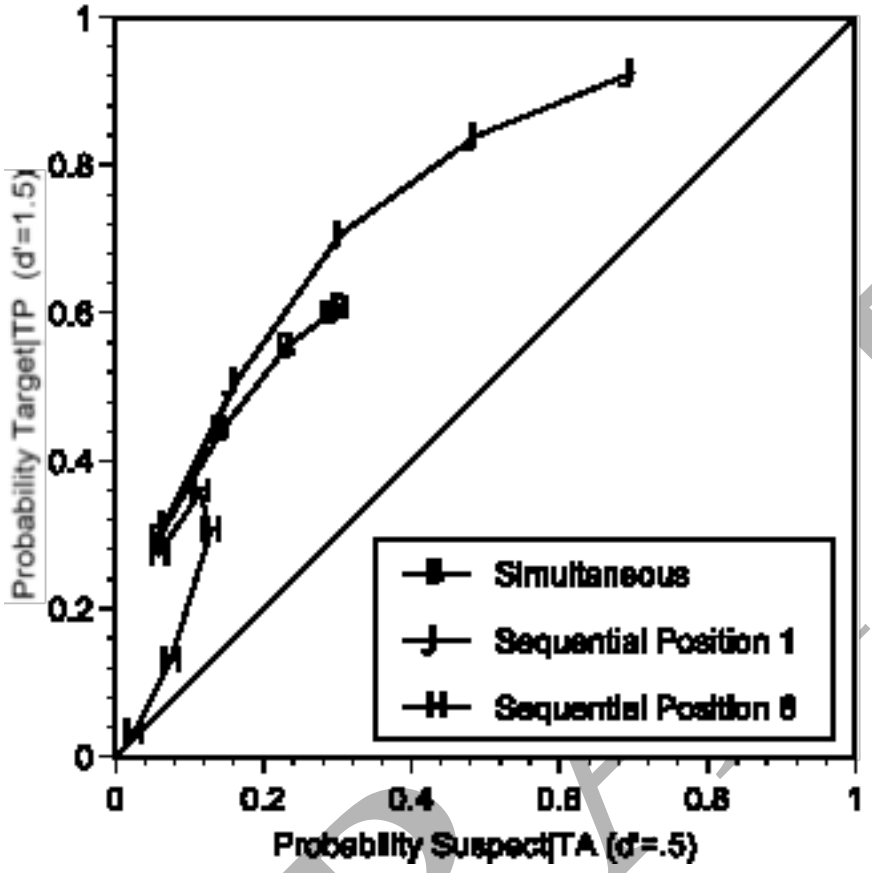
DRAFT


Probability Hit|Target Present
Probability False Alarm|Target Absent
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0
0.2
0.4
0.6
0.8
1

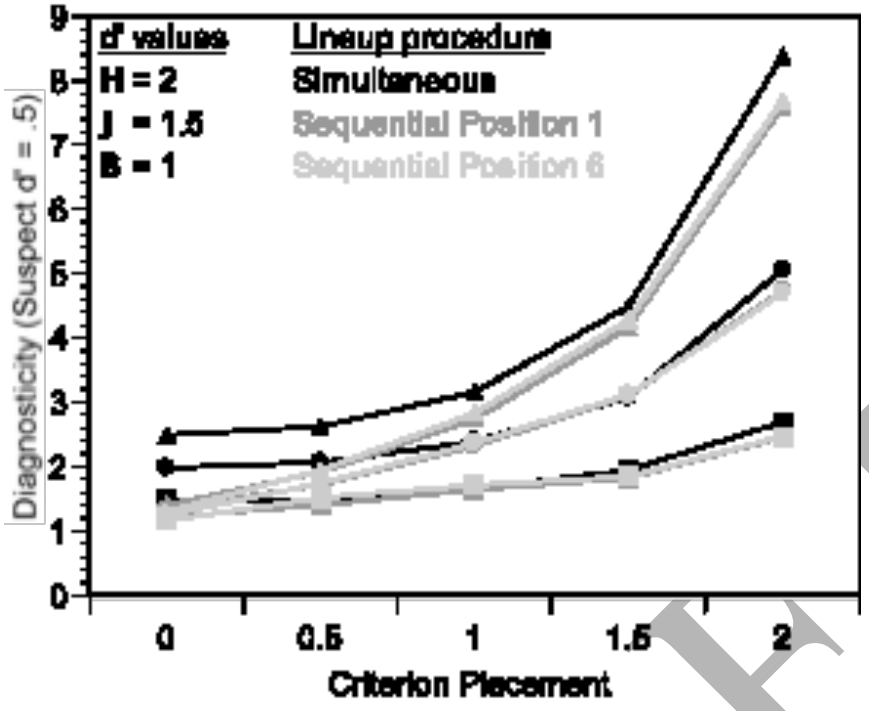
DRAFT


B d' = 2
J d' = 1
H d' = 0
Sequential Position 1
Sequential Position 6
Simultaneous
Prob Hit|Target Present Lineup
Prob FA|Target Absent Lineup
0
0.2
0.4
0.6
0.8
1

DRAFT


Probability Target|TP (d'=1.5)
Probability Suspect|TA (d'=.5)
0
0.2
0.4
0.6
0.8
1
Simultaneous
Sequential Position 1
Sequential Position 6

DRAFT



DRAFT

Behrman, B. W., & Davey, S. L. (1999). Eyewitness memory for actual crimes: An archival analysis. Paper presented at the American Psychological Society, 11th Conference, Denver, CO.

Behrman, B. W., & Vayder, L. T. (1994). The biasing influence of a police showup: Does the observation of a single suspect taint later identification? Perceptual & Motor Skills, 79(3, Pt 1), 1239-1248.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. Cognitive Science, 23(1), 1-31.

Corey, D., Malpass, R. S., & McQuiston, D. E. (1999). Parallelism in eyewitness and mock witness identification. Applied Cognitive Psychology, 13(Spec Issue), S41-S58.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? The view from UCSD. In M. J. Wenger & J. T. Townsend (Eds.), Computational, geometric, and process perspectives on facial cognition (pp. 347-395). Mahwah, NJ: Lawrence Erlbaum Associates.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. Journal of Applied Psychology, 73(2), 281-290.

Ebbesen, E. B., & Wixted, J. (1996). A signal detection analysis of the relationship between confidence and accuracy in eyewitness memory. La Jolla: University of California, San Diego.

Eyewitness evidence: A guide for law enforcement. (Research report)(1999). Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.

Farah, M. J. (1996). Is face recognition "special"? Evidence from neuropsychology. Behavioural Brain Research, 76(1-2), 181-189.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? Psychological Review, 105(3), 482-498.

Flowe, H., Ebbesen, E. B., Burke, C., & Chivabunditt, P. (2001). At the scene of the crime: An examination of the external validity of published studies on line-up identification accuracy. Toronto, Canada: American Psychology Society Annual Convention.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. Journal of Personality & Social Psychology, 64(4), 525-537.

Hancock, P. J. B., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. Memory & Cognition, 24(1), 26-40.

Levi, A. M. (1998). Protecting innocent defendants, nailing the guilty: A modified sequential lineup. Applied Cognitive Psychology, 12(3), 265-275.

Lindsay, R. C., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. Journal of Applied Psychology, 76(5), 741-745.

Lindsay, R. C., Lea, J. A., Nosworthy, G. J., & Fulford, J. A. (1991). Biased lineups: Sequential presentation reduces the problem. Journal of Applied Psychology, 76(6), 796-802.

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. Journal of Applied Psychology, 70(3), 556-564.

Lindsay, R. C. L. (1999). Applying applied research: Selling the sequential line-up. Applied Cognitive Psychology, 13(3), 219-225.

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. Journal of Applied Psychology, 84(3).

Lindsay, R. C. L., Craig, W., Lee, K., Pozzulo, J. D., Rombough, V., & Smyth, L. (1995, July). Eyewitness identification procedures for use with children. Paper presented at the Society for Applied Research in Memory and Cognition, Vancouver, British Columbia.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., & Lee, K. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. Law & Human Behavior, 21(4), 391-404.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York, NY, USA: Cambridge University Press.

Melara, R. D., DeWitt-Rickards, T. S., & O'Brien, T. P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. Journal of Applied Psychology, 74(5), 706-713.

Navon, D. (1990a). Ecological parameters € nonlineup evidence: A reply to Wells and Luus. Journal of Applied Psychology, 75(5), 517-520.

Navon, D. (1990b). How critical is the accuracy of an eyewitness's memory? Another look at the issue of lineup diagnosticity. Journal of Applied Psychology, 75(5), 506-510.

DRAFT

Navon, D. (1991). "Ecological parameters € nonlineup evidence: A reply to Wells and Luus": Correction. Journal of Applied Psychology, 76(3), 407.

Navon, D. (1992). Selection of lineup foils by similarity to the suspect is likely to misfire. Law & Human Behavior, 16(5), 575-593.

O'Toole, A. J., Wenger, M. J., & Townsend, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In E. Michael J. Wenger & E. James T. Townsend (Eds.), Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. (pp. 1-38): Mahwah, NJ, US.

Parker, J. F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. Law & Human Behavior, 17(1), 11-26.

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. Journal of Applied Psychology, 84(2).

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. Psychological Review, 99, 518-535.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. Psychological Review, 108(2).

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. Journal of Applied Psychology, 78(1), 22-33.

Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In J. D. R. M. P. T. E. David Frank Ross (Ed.), Adult eyewitness testimony: Current trends and developments. (pp. 300-327): Cambridge University Press, New York, NY, US.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. Law & Human Behavior, 21(3), 283-297.

Tollerstrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In J. D. R. M. P. T. David Frank Ross (Ed.), Adult eyewitness testimony: Current trends and developments. (pp. 144-160): Cambridge University Press, New York, NY, US.

Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? Law & Human Behavior, 24(2), 231-258.

Wells, G. L. (1984). The psychology of lineup identifications. Journal of Applied Social Psychology, 14(2), 89-103.

Wells, G. L. (1993). What do we know about eyewitness identification? American Psychologist, 48(5), 553-571.

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. Psychological Bulletin, 88(3), 776-784.

Wells, G. L., & Luus, C. E. (1990). The diagnosticity of a lineup should not be confused with the diagnostic value of nonlineup evidence. Journal of Applied Psychology, 75(5), 511-516.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. Journal of Applied Psychology, 78(5), 835-844.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. Law & Human Behavior, 22(6).

DRAFT

Yarmey, A. D. (1998). Person identification in showups and lineups. In E. Charles P. Thompson & E. Douglas J. Herrmann & et al. (Eds.), Eyewitness memory: Theoretical and applied perspectives. (pp. 131-154): Mahwah, NJ, USA.

Yarmey, A. D., & Morris, S. (1998). The effects of discussion on eyewitness memory. Journal of Applied Social Psychology, 28(17), 1637-1648.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identification in showups and lineups. Law & Human Behavior, 20(4), 459-477.

DRAFT