Title
The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups

Author
Flowe, Heather D.

Publication Date
2005

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

The Effect of Lineup Member Similarity on Recognition Accuracy in Simultaneous and Sequential Lineups

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy in

Psychology

by

Heather D. Flowe

Committee in Charge:

Professor Ebbe B. Ebbesen, Chair
Professor Nicholas Christenfeld
Professor Garrison W. Cottrell
Professor Hugh Mehan
Professor John T. Wixted

2005

The dissertation of Heather D. Flowe is approved, and it is acceptable in quality and form for publication on microfilm.

____

____

____

____

____

Chair

University of California, San Diego

2005

DEDICATION

In recognition of their love and support, this thesis is dedicated to my family, especially to my father, Butch Flowe.

TABLE OF CONTENTS

List of Tables

List of Figures

VITA

August 16, 1973	Born, Cocoa Beach, Florida
1997	B.A., University of California, San Diego
1998-1999	Research Associate, Department of Anesthesiology, University of Michigan, Ann Arbor
2000	M.A., University of California, San Diego
2000-2003	National Science Foundation Graduate Fellow
2002-2005	Adjunct Faculty Member, Grossmont College, El Cajon, California
2003	Summer Graduate Teaching Fellow, University of California, San Diego
2003-present	Instructor, Department of Psychology, University of California, San Diego
2005-present	Adjunct Faculty Member, San Diego City College, San Diego, California
2005	Ph.D., University of California, San Diego
2005	Faculty Fellow, University of California, San Diego

Publications:

Green, C. R., Flowe-Valencia, H. D., Rosenblum, L., & Tait, A. R. (2001). The role of childhood and adulthood abuse among women presenting for chronic pain management. Clinical Journal of Pain, 17, 359-364.

Haefner, H. K., Khoshnevisan, M. H., Bachman, J. E., Flowe-Valencia, H. D., Dent, D. M. D., Green, C. R., & Reed, B. D. (2000). Use of the McGill Pain Questionnaire to compare women with vulvar pain, pelvic pain and headaches. Journal of Reproductive Medicine, 45, 665-671.

Green, C. R, Flowe-Valencia, H. D., Rosenblum, L., & Tait, A. R. (1999). Do physical and sexual abuse differentially affect chronic pain states in women? Journal of Pain and Symptom Management, 18, 420-426.

Professional Presentations:

Flowe, H. D. (2004). A short history of eyewitness identification procedures. Paper presented at the University of California San Diego Graduate Student Symposium, La Jolla, California.

Flowe, H. D. & Ebbesen, E. B. (2003). The misinformation effect: Trace Alteration or Coexistence? Paper Presented at the Society for Applied Research in Memory and Cognition, Aberdeen, Scotland.

Flowe, H. D., Ebbesen, E. B. (2003). Simultaneous v. sequential lineups: What do we know? Paper presented at the American Psychology Law Society, Edinburgh, Scotland.

Flowe, H. D. (2002). An analysis of mistaken identification claims. Paper presented at the University of California San Diego Graduate Student Symposium, La Jolla, California.

Flowe, H. D., & Ebbesen, E. B. (2002). An archival analysis of trial cases in which the defendant claimed mistaken identification. Paper presented at the American Psychology and Law Society, Austin, Texas.

Flowe, H. D., & Ebbesen, E. B. (2001). Routinely measuring lineup member similarity in experimental research is a good idea. Paper presented at the Society for Applied Research in Memory and Cognition, Kingston, Canada.

Flowe, H. D., & Ebbesen, E. B. (2001). At the scene of the crime: An examination of the external validity of published studies on line-up identification accuracy. Poster presented at the American Psychological Society, Toronto, Canada.

Flowe, H. D. (2000). Mistakenly Identified: An examination of bias in a random sample of criminal line-ups. Paper presented at the University of California San Diego Graduate Student Symposium, La Jolla, California.

Flowe, H. D., & Ebbesen, E. B. (2000). "Nice girls don't get raped and bad girls shouldn't complain." The relationship between sexual behavior and rape accusations. Paper presented at the American Psychology and Law Society, New Orleans, Louisiana.

Baugh R., Green, C. R., & Flowe-Valencia, H. D. (1998). A Retrospective Review of Elderly Patients Treated with Epidural Analgesia for Postoperative Pain Control. Paper presented at the Midwest Anesthesia Residents Conference, Minneapolis, Minnesota.

LaGorio J, Green, C. R., & Flowe-Valencia, H. D. (1998). Thoracic Epidural Analgesia Compared to Other Methods of Analgesia for Single Lung Transplant. Paper presented at the Midwest Anesthesia Residents Conference, Minneapolis, Minnesota.

Minott S., Green, C. R., & Flowe-Valencia, H. D. (1998). Evaluation of Naloxone Administration to Patients on the Acute Pain Service. Paper presented at the Midwest Anesthesia Residents Conference, Minneapolis, MN. 1998.

University Service:

2003-2004	Member of the UC Office of the President Committee on Credit Cards on Campus
2002-2004	Board Member of the University of California Student Assn.
2002-2004	Vice President, External Affairs, UCSD Graduate Student Assn.
1999-2002	Psychology Dept. Rep. to the UCSD Graduate Student Assn.
1999-2000	Member, UC San Diego Academic Senate Program Review Committee

ABSTRACT OF THE DISSERATATION

The Effect of Lineup Member Similarity on Recognition Accuracy in

Simultaneous and Sequential Lineups

by

Heather D. Flowe
Doctor of Philosophy in Psychology
University of California, San Diego, 2005

Professor Ebbe B. Ebbesen, Chair

Prior research suggests that simultaneous lineup procedures, which involve presenting at once all of the alternatives, encourage witnesses to identify the relatively most familiar face. In contrast, when lineup alternatives are evaluated one at a time, or sequentially, identifying a face is not influenced by previously seen faces in the lineup. Instead, a face will be chosen from a sequential lineup only when it is a sufficient match to the perpetrator in memory. The present study investigated whether face discrimination differs between simultaneous and sequential lineups. In particular, if the retrieval process is influenced by comparing the faces during the decision process, then the similarity of the alternatives with respect to the perpetrator should influence accuracy only in simultaneous lineups. Overall, positive identifications of the target, whether identical to the study face or a feature substituted version of the study face, were higher in simultaneous lineups, while misses occurred with greater frequency in sequential lineups. We also found that in both simultaneous and sequential presentations, decreasing the similarity of the alternatives increased the rate at which the innocent suspect was identified. In addition, if an alternative that was similar to the study face was presented before the target, sequential participants tended to identify it and miss the target face as a result. When the target was removed from the lineup, there was a tendency in both simultaneous and sequential conditions to choose the alternative that was the highest in similarity with respect to the target. Accuracy was correlated with lineup fairness and similarity measures, and the pattern of findings was similar across identification procedures. A final set of studies examined whether choice rates in sequential presentations are lower because participants have a higher identification threshold than simultaneous participants. To investigate, the number of features that varied across the alternatives was manipulated, in addition to whether participants were forced to choose a face or given the option not to choose. Accuracy was higher in simultaneous compared to sequential lineups when fewer features were available for discriminating between the faces in the lineup. However, under forced choice conditions, differences in accuracy between identification procedures were significantly decreased, suggesting that not being able to compare the faces might lead witnesses to adopt a higher decision criterion. The findings from the series of studies reported are evaluated with respect to theoretical models of eyewitness identification.

INTRODUCTION

Much psychological research over the last 25 years has demonstrated the conditions under which eyewitnesses might be prone to making identification errors. Based on what we have learned from these studies, psychologists have called on the legal system to make procedural changes aimed at reducing the rate of mistaken identification (Wells et al., 1998). In particular, psychologists have begun to recommend the use of the sequential lineup over the traditional simultaneous lineup (e.g., Lindsay et al., 1991; Wells et al. 1998). Across studies it has been shown that by switching to sequential lineups, false identifications are reduced, while positive identifications, under laboratory conditions that might approximate "real world" conditions, are largely unaffected (see Steblay et al., 2001 for a meta-analytic comparison of simultaneous and sequential lineups).

In a simultaneous lineup, witnesses are shown at once an array of faces, which includes the suspect along with persons known by the police to be innocent (foils). In a sequential lineup, faces are viewed one at a time, and for each face the witness makes a yes/no decision. The next picture is displayed if the photo is rejected by the witness, and once a photo has been rejected, the witness is not allowed to see it again. The procedure continues until a face is positively identified as the culprit. No other photos are presented once a photo has been identified.

In the seminal study comparing the identification outcomes from the two procedures, Lindsay and Wells (1985) hypothesized that simultaneously displaying the photographs encourages witnesses to choose the person in the lineup who looks relatively the most similar to the suspect. They further hypothesized that presenting the faces sequentially would prevent witnesses from making relative comparisons among lineup members. Witnesses who see the faces sequentially, therefore, will base their judgment on their memory for the culprit, which they termed as an absolute judgment strategy, rather than on relative similarity. Lindsay and Wells (1985) found no difference between the two identification procedures in hits when the perpetrator was present in the lineup (sequential M=.50 and simultaneous M=.58); but, a sizeable difference in false identifications of the innocent look-a-like was obtained (sequential M=.17 and simultaneous M=.43) when the perpetrator was absent from the lineup, thereby leading them to conclude that their hypotheses had been supported.

Subsequent research has replicated Lindsay and Wells (1985), finding that false alarms are reduced while hit rates are not appreciably affected (Cutler & Penrod, 1988; Levi, 1998; Lindsay, et al., 1991; Lindsay & Wells, 1985; Lindsay, 1999; Lindsay, Pozzulo, Craig, & Lee, 1997; Parker & Ryan, 1993; Sporer, 1993). Since it seems as though the procedure will protect innocent suspects while still allowing for the identification of the guilty, psychologists have recommended that police investigators conduct lineups sequentially. In an official position paper, for instance, the American Psychology Law Society argued that sequential lineups protect innocent suspects while allowing for the conviction of the guilty (Wells et al., 1998). The paper also maintains that sufficient research evidence exists to support the idea that witnesses are prone to making relative judgments in simultaneous lineups. The arguments put forth were instrumental in leading New Jersey to become the first state to adopt exclusively the sequential lineup procedure. In advocating for the use of sequential lineups, Attorney

General John Farmer 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" (Farmer, 2001). However, as we will show in the next section, the research evidence supporting the idea that a relative judgment process underlies decision making in simultaneous lineups is rather indirect.

We undertook the current project to examine in a more direct manner whether the face recognition process is different in simultaneous compared to sequential lineups. The paper is organized as follows: First, we describe the evidence supporting the hypothesis that different recognition processes occur depending on whether the test is administered sequentially or simultaneously. Second, the models that have been put forth to describe face recognition in lineups are reviewed. Third, we report results from a series of studies that examined the impact of lineup member similarity on identification accuracy in simultaneous and sequential lineups. If participants compare alternatives when the lineup is presented, then the similarity of the alternatives with respect to the suspect should have little consequence for whether or not he is identified. On the other hand, if alternatives are compared and the relatively best alternative is chosen, then the foils should have more of an effect. Based on this line of reasoning, similarity manipulations should affect accuracy in only simultaneous lineups. The last section discusses how well the results from the original research we report fit with existing models of eyewitness identification, as well as with other theoretical developments regarding the impact of lineup similarity structure on identification accuracy. Limitations of the methods that we used and the directions that future research and theoretical development might take are also considered.

Evidence of Absolute versus Relative Judgments

Two basic judgment processes are thought to underlie identification from lineups (Wells et al., 1998). A relative judgment process entails selecting the face in the lineup that is the best match to the study face/culprit in memory relative to the others. An absolute strategy involves selecting a face from the lineup that is the best match to the culprit in memory. Both judgment processes involve memory comparisons; what differs primarily between the two processes is whether the foils play a part in the recognition process. In the relative process the foils influence recognition, whereas in the absolute process they do not.

Wells et al. (1998) cited the removal without replacement procedure as the best evidence that eyewitnesses use a relative judgment strategy in simultaneous lineups. The removal without replacement testing paradigm involves two parts: After viewing the to-be-identified target, participants are shown a culprit-present simultaneous lineup, and the distribution of lineup responses is recorded. In the second part, another group of participants undergoes similar experimental procedures, except that the culprit has been removed from the lineup and not replaced by any another face. If responses are based on relative judgments, then the most widely chosen foil from the full lineup should be chosen from the removed lineup at an even higher rate. On the other hand, if witnesses were using an absolute judgment strategy, they would correctly reject the removed lineup because the culprit is not there. In particular, the lineup should be rejected at a rate equal to the proportion rejecting the lineup and the proportion selecting the target in part one. Wells (1993) found that after the culprit had been removed, the rate at which the most popular foil was chosen in the full lineup was significantly increased in the removed lineup. Thus, it appears when the target is removed from the lineup, simultaneous participants shift their choice to the next best alternative.

Recently, however, Clark and Davey (2005) demonstrated in sequential lineups that the foil identified most often in a full lineup also draws a larger number of choices when the culprit is removed. Moreover, they found position effects in the full lineup such that when the most popular foil was presented before the target, sequential participants tended to miss the target because they chose the earlier presented foil instead. The position of the most popular foil in the target removed lineup did not affect the rate at which the popular foil was chosen. In a second experiment, they reduced the similarity of the most popular foil as well as the similarity of the other foils in the lineup relative to the target. Once again, the most popular foil was chosen at a higher rate in the target removed lineup both the simultaneous and sequential conditions. In addition, participants chose the most popular foil in removed lineups at a higher rate if he was in position 4 compared to position 2. A within lineup criterion shift was proposed to account for the findings. Specifically, when the next best alternative was presented early in the lineup, participants withheld making a choice in order to find out whether a better option would be presented later in the lineup. Only low similarity foils appeared thereafter, and consequently participants rejected the lineup. In contrast, after seeing low similarity foils, if the next best alternative was presented later in position 4, participants lowered their decision criterion and chose him. There are several reasons, however, to remain cautious about the idea that participants were shifting their decision criterion face by face during the task. First, previous research in the false memory domain suggests that participants do not shift their decision criterion item by item during a memory task, even under conditions in which they should know that it would be advantageous to do so (Stretch & Wixted, 1998; Wixted & Stretch, 2000). Second, if participants lowered their criterion during the task, the rate of choosing the target should have been higher in later compared to earlier positions, but no such position effects for the target were obtained. Third, a memory based explanation can also account for the results, an issue that will be explored in greater detail in subsequent sections. For now, perhaps a face within the lineup that does not strongly match the contents of memory might be chosen if it has relatively greater associative strength with the culprit than the low similarity faces that were presented earlier. In any case, any memory model of the findings would have to simultaneously account for why the target was not chosen at a higher rate in position 4 compared to position 2, and why position effects for the popular foil in the removed lineups were found only when the similarity of both the popular foil and the other foils relative to the target were reduced. Putting aside the issue of the explanation behind the increase in choosing the next best foil, the Clark and Davey (2005) findings illustrate that in sequential lineups, the probability of selecting a given face is influenced by earlier presented faces. Therefore, memory models of identification in sequential lineups may have to incorporate foil parameters.

A second source of evidence that relative judgments are made in sequential 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 simultaneous lineup are less apt to make false identifications than eyewitnesses who are not given such instructions. Steblay (1997) found in a meta-analysis that admonishment to use a stricter criterion only seems to affect errors made to target absent lineups. This outcome has been taken to mean that admonishment against erroneous identifications is an indirect means by which to discourage relative decision making in lineups, thereby increasing the extent to which decisions are made based on match-memory evidence instead (Wells et al., 1998). In addition, Lindsay et al. (1991) found evidence that suggests that criterion manipulations (i.e., biased instructions that indicate that the target is present compared to fair instructions that indicate that the guilty party might not be in the lineup) in target absent lineups may have a larger effect on accuracy in simultaneous compared to sequential lineups. Specifically, correct rejections increased from 33% in the biased condition to 60% in the fair condition in simultaneous lineups, and from 77% to 87% in sequential lineups. These findings suggest that sequential participants might be more biased against positively identifying any face than simultaneous participants.

Another research finding cited in support of the idea that relative judgments are made in simultaneous lineups is the effect that the dual lineup procedure has on accuracy rates (Wells et al., 1998). In the dual lineup procedure, eyewitnesses are first shown a blank lineup (i.e., a lineup that does not contain the suspect) before 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 colleagues (1998) argued that blank lineups might be used to screen out witnesses who are prone to making relative judgments.

Lastly, in addition to the experimental research, self-report data are consistent with the idea that participants make absolute judgments in sequential lineups and relative judgments in simultaneous lineups (Kneller, Memon, & Stevenage, 2001; Lindsay & Bellinger, 1999; Lindsay, Lea, Nosworthy, 1991; Smith, Lindsay, & Pryke, 2000). 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 are more accurate if they report using an absolute rather than a relative strategy. Some subjects, however, have been known to report using an absolute strategy even though the experimenter observed them comparing lineup pictures (Lindsay & Bellinger, 1999). Equally important, the factors that control self-reports of mental processes might not affect the actual decision processes. For example, subjects might report more often that they compare pictures in a simultaneous lineup because they can shift their visual gaze from one picture to another and cannot do so in a sequential lineup. Gaze shifting might have little to do with whether subjects are using an absolute similarity standard. Additionally, relative comparisons might also be made in sequential lineups, with witnesses mentally comparing each photo to previously seen ones.

<u>Other Models of Face Recognition in Lineups</u>

We (Ebbesen & Flowe, 2001) cast simultaneous and sequential lineup identifications within a signal detection framework to better understand how false alarms are affected to a greater extent than positive identifications of the culprit. Specifically, in addition to differences between the two procedures in false identifications of the innocent suspect, or in false alarm rates, small but nevertheless consistent differences in positive identifications of the culprit, or on the hit rate, occur between the two procedures. Across studies, sequential participants are about 15% less likely to positively identify the culprit (Steblay et al., 2001). Moreover, we questioned the extent to which differential reliance on relative decision making accounted for the difference in results between the two procedures. If witnesses were simply selecting the relatively most familiar person, then they should almost always select someone. However, participants in the laboratory (Steblay et al., 2001) and witnesses in real world cases (Behrman & Davey, 2001) more than 50% of the time do not identify anyone from a simultaneous lineup.

Differential decision criterion placement was proposed to account for the differences in identification outcomes found between the two lineup procedures, and it was based on the following reasoning: Besides seeing only one picture at a time, an additional feature that differentiates sequential lineup presentations is that witnesses do not know the number of photographs they will be seeing. Instead, a stack of photographs is displayed to give the impression that many will be shown. This procedural detail was added in order to prevent people from picking the last photograph simply because they know no others will be displayed. The identification session ends after a predetermined number of pictures (usually 6) have been presented. One potential effect of this procedure, however, is that witnesses will withhold making a choice, concerned that a better match to the culprit is still to come. In other words, witnesses viewing sequential lineups might set their decision criteria higher than those viewing simultaneous lineups might, and as a consequence, reject the lineup at a higher rate. Thus, higher criterion placement would account for the decrease in hits and false alarms found for sequential lineups.

A signal detection analysis can also account for the differential effects on hits and false alarms that have been observed in comparing the two lineup procedures (see Figure 1). The underlying model of signal detection theory consists of two normal distributions, one representing the signal and the other noise. In the current application of signal detection theory, the signal distribution is the culprit distribution, and the foil (and the innocent suspect look-a-like if the culprit is not in the lineup) distributions are the noise distributions. In most signal detection applications, the distributions are aligned on a strength of memory to evidence dimension. For the present purpose, the dimension of interest is familiarity, which is conceived of as the lineup member's similarity to the culprit in memory. If switching to sequential lineups causes witnesses to place their criterion at a higher level, then witnesses will be less likely to choose an innocent suspect if the pictures are presented one at a time (see bottom panel of Figure 1). Furthermore, as Figure 1 demonstrates, the effect on false alarms will be greater than the effect on hit rates as a result of switching procedures when 1) a memory of the culprit was laid down, 2) the target is on average more familiar than the foils, and 3) if sequential participants are more biased to reject rather than identify a face compared to simultaneous participants. In addition, the effect on hits might even be larger than that depicted in Figure 1, as learning might not only increase the mean of the target distribution, but also increase the variance (Ebbesen & Wixted, 1996; Ratcliff, Sheu, & Gronlund, 1992).

Grolund (2004) made a comparison of the relative/absolute and criterion accounts of lineup identification by presenting witnesses with men of different heights who were either encoded in a relative ("he is taller than") or an absolute manner ("he is 5'7""). At test, each man was presented in a lineup, with each picture in the lineup being of the same man shown at different heights. He expected that accuracy would be higher under test conditions that matched study conditions. He found that discriminability was higher in simultaneous lineups if the heights were originally learned under relative compared to absolute conditions, while in sequential lineups, discriminability was higher if the men were learned under absolute compared to relative conditions. Based on this pattern, he concluded that his findings were consistent with the absolute and relative decision models of eyewitness accuracy, not the criterion shift model. How well his study procedures capture face recognition processes in general and lineup recognition in particular is open to interpretation however. First, the extent to which these findings involving a global feature such as height apply to face stimuli—including the first order features and the spatial relations among them, is in need of demonstration. For instance, distinguishing faces on the basis of some features, such as height and weight, might naturally lend themselves well to relative comparisons because there is a more obvious central tendency in the population for those features, while others, such as the distance between the eyes or the width of the nose have a less obvious standard by which comparisons can be made. Second, there were no significant differences overall (collapsing across encoding condition) in hits or false alarms between simultaneous and sequential lineups. This pattern is different from that which is usually observed experimentally between the two procedures (i.e., bigger effect on false alarms than on hits). Thus, while this test of the criterion shift and relative/absolute decision models is innovative, it remains to be seen whether the different patterns found between the two identification procedures emerge for face recognition in lineups.

More recently, the WITNESS model of lineup identification has been proposed, and it reflects the ideas that identification is a weighted combination of absolute and relative judgments along with criteria for identification and lineup rejection (Clark, 2003). For instance, if a lineup face is compared to the memory trace of the culprit and the match to memory is very high, the face might be selected without consideration of the similarity between the chosen face and any of the foils. Another means by which a face can be selected is strictly through a relative judgment strategy. The model instantiates the relative judgment process by taking the difference between the two faces that have the best and the next best similarity match with the culprit in memory. If the difference is high, the best face is selected. The decision could also be based on a combination of the absolute and relative strategies, wherein the best match and the best-next difference are weighted together, and if their conjunction is higher than the identification criterion value, the face is selected. The decision to reject the lineup is based on an absolute matching strategy alone, such that if all of the faces are below the criterion value for rejection then the lineup is rejected. If all of the faces do not fall below the rejection criterion, then a “don’t know” response is given instead of a rejection. In addition, the WITNESS model includes similarity parameters between the innocent suspect-to-perpetrator, the foils-to-suspect, and the foils-to-perpetrator, which are used in the matching process. If the foils bear little resemblance to the perpetrator, while the suspect in the lineup is a strong match to the perpetrator (either because he is the culprit or is a dead ringer), the suspect may be identified at a rate significantly above chance levels, even under conditions in which memory for the culprit is weak, as the difference between the best and the next best lineup member will tend to be large. However, if the suspect is high in resemblance to both the foils and the perpetrator, then a "don't know" response is likelier than a positive identification.

SUSPECTS, proposed by Grolund (2005) as another candidate model for explaining differences between the two identification procedures, proposes that the process of recollection is more likely in sequential than in simultaneous lineups. As a result, witnesses viewing sequential lineups are more likely retrieve distinctive information about the perpetrator. Distinctive information, in turn, plays a more decisive role in rejecting rather than in identifying a face. That is, distinctive information is used to rule out alternatives. Grolund (2005) hypothesizes recollection is less likely in simultaneous lineups because it requires the witness to attend to many faces at once. Consequently, less attention remains for considering distinctive features. Research suggests, however, that simultaneous witnesses do attend to distinctive features (e.g., Brigham, 1990; Brigham et al., 1982). Moreover, the SUSPECTS model would need to account for why distinctiveness does not play a role in correct identifications. Specifically, the hit rate in simultaneous lineups is just as high, if not higher, than sequential lineups, which demonstrates that simultaneous participants are just as able to correctly reject (discriminate) incorrect alternatives. In addition, simultaneous witnesses appear to spend an equal amount of time processing the faces in the lineup as do sequential participants, suggesting that attention allocated to individual faces between the two procedures might be similar (Kneller, Memon, & Stevenage, 2001; Sporer, 1993).

The Effects of Lineup Member Similarity in Simultaneous and Sequential Lineups

In the series of experiments that follow, we set out to test whether the recognition process differs depending on whether lineup alternatives are presented simultaneously or sequentially. The point of these experiments was not so much to determine which of the foregoing models of eyewitness identification is correct, but rather to address the specific question of whether the foils play a greater role in recognition from simultaneous compared to sequential lineups. If they do, then this suggests that the recognition process is different depending on the lineup procedure employed. Before describing the way in which lineup member similarity was operationalized and our predictions, findings regarding the impact of lineup member similarity are briefly reviewed next.

Overview of Empirical Findings on Lineup Member Similarity and Accuracy. Most of the empirical work on lineup member similarity so far has been conducted on simultaneous lineups. In general, it appears that the less similar the foils are to the perpetrator, the more likely it is that the suspect will be identified (see Navon 1992, for mathematical model of how similarity relations among the members and the suspect can influence accuracy). For instance, empirically it has been demonstrated that if the foils are chosen for a lineup based on their match in physical appearance to the innocent suspect rather than to the culprit, false identifications of the innocent suspect are increased (Clark & Tunnicliff, 2001). Since the foils were chosen for the lineup based on their match to the innocent suspect, and the suspect was chosen ostensibly based on his match to the perpetrator, the suspect will look most like the perpetrator.

Another way in which similarity of the foils with respect to the target can be manipulated is by matching the foils to a description of the perpetrator or by matching the foils to the physical appearance of the suspect (Luus & Wells, 1993). If the foils are matched to an innocent suspect, then similarity among the members with respect to the culprit is reduced. The findings with respect to match-suspect versus match-description on identification accuracy, however, are mixed (see Clark, 2003 and Tunnicliff & Clark, 2001 for an analysis of the differences among studies). For instance, Luus and Wells (1993) were the first to hypothesize that identification accuracy might be influenced depending on whether the foils were chosen based on their match to the suspect or their match to the witness' description of the perpetrator. Wells, Rydell, and Seelau (1993) performed an experiment to test these predictions, and while they found no differences in innocent suspect identifications, hits where significantly decreased when the lineup was suspect-matched rather than description-matched. These findings suggest that the foils were too similar to the target in the suspect-matched condition and thus discrimination was poor.

Lindsay, Martin and Webber (1994) also found an effect on hits that favored description-matched over suspect-matched lineups; however, they found that false alarms were higher in description-matched compared to suspect-matched lineups under conditions in which the foils fit a vague description of the perpetrator and were otherwise maximally dissimilar to the perpetrator. Interestingly, they ran a sequential condition and found, despite the fact that choice rates were substantially lower in sequential compared to simultaneous lineups, false alarms to the innocent suspect were also increased in sequential lineups when the foils were description-matched as opposed to suspect-matched. In addition, participants did not seem to be selecting the lure because he was a dead ringer for the perpetrator, as experimenters chose to use a lure that was moderate in similarity to the perpetrator based on similarity ratings.

More recently, in an attempt to discover whether the match-description advantage found by Wells et al. (1993) could be obtained for lineups created by police officers, Clark and Tunnicliff (2001) found in two experiments that hits and false alarms did not differ for suspect- and description-matched lineups, though similar to Lindsay et al. (1991), they found fewer correct rejections when the lineup was description-matched compared to suspect-matched. The mixed results across studies may be attributable to differences across studies in the way in which similarity was controlled across foils and across lineup members. Clark and Tunnicliff (2001) speculate that Luus and Wells (1993) might have utilized foils that were on average more similar to the perpetrator.

Differences across studies in the degree to which the innocent suspect resembled the culprit may also have influenced the results. In both Wells et al. (1993) and Tunnicliff and Clark (2001), the innocent suspects differed from lineup to lineup because they were chosen from a pool based on the witness' description. In contrast, Lindsay et al. (1994) chose a moderately similar lure based on direct similarity comparisons with the perpetrator. Among 199 faces, the lure was placed in the top quartile with respect to his similarity with the perpetrator, and against a backdrop of description-matched foils, this familiarity of the suspect might have been further enhanced. Therefore, perhaps false alarms to the innocent suspect were higher in Lindsay et al. (1994) because the innocent suspect was more similar to the perpetrator than in the other two studies.

Operationalization of Similarity in the Present Study. Taken together, the mixed results from the description-matched versus suspect-matched literature illustrate that there are several similarity dimensions within a lineup that might influence accuracy. In the present study, all similarity dimensions, including foil-culprit, foil-suspect, and foil-foil were taken into consideration. In addition, composite drawings of faces were utilized as face stimuli so as to tightly control similarity.

The similarity of the foils relative to the suspect was systematically controlled by selecting foils that matched the perpetrator based on a single feature, or selecting foils that matched the perpetrator based on a general description of the target, but otherwise had no facial features in common with the perpetrator (i.e., the foils were randomly chosen). Based on this methodology, the similarity of the foils relative to the perpetrator on average should have been more similar in the matched condition compared to the random condition. We opted to match the foil to the perpetrator rather than to the suspect so as to maximize foil choices in order to determine whether discrimination between the foils and the target was greater in sequential compared to simultaneous lineups. That is, if lower match to memory evidence is utilized when all of the alternatives are viewed at once rather than one at a time, then foil selections should be higher in simultaneous compared to sequential lineups when there are foils present that match the perpetrator on a particular feature.

Second, we systematically varied the similarity of the suspect relative to the study face by using the feature substitution method. Specifically, one feature from the study face was removed and replaced by another feature belonging to a different feature class (e.g., if the study face had bulging eyes, the substituted feature was almond shaped eyes). In so doing, we were able to determine whether sequential participants were better able to discriminate subtle changes than simultaneous participants, a hypothesis that is in keeping with the idea that sequential participants use an absolute judgment strategy.

Overview of Predictions and Experiments. In the series of experiments that follow, the following questions were of main concern:

1) Is target discrimination affected by the similarity of the alternatives in the lineup?

If identifications from simultaneous lineups are based on relative comparisons, then surrounding the target with alternatives that are dissimilar to the perpetrator should increase the rate at which the target (whether "guilty" or "innocent") is chosen. In contrast, if absolute judgments are made in sequential lineups, the similarity of the foils will not influence accuracy. Experiments 1-3 tested these hypotheses.

2) How do variables that affect decision criterion placement affect accuracy in simultaneous and sequential lineups?

If initial criterion placement differs between the two procedures, then increasing the decision criterion should have a bigger effect on hits than false alarms in sequential lineups, and a bigger effect on false alarms than hits in simultaneous lineups, predictions that were tested in Experiment 1 and 5.

3) Do participants focus on different featural aspects of faces depending on whether they view all of the alternatives together or separately?

Prior research has found that internal features of faces are more salient than external features, and that the upper part of the face is more salient than the lower part of the face (Shepherd, Davies, and Ellis, 1981). In addition, other work suggests that reliance on specific features might be important depending on the task at hand, such as emotion discrimination (Cottrell, Dailey, Padgett, & Adolphs, 2001). If discrimination is better in sequential lineups, perhaps it is because features, whether high or low in salience, are analyzed in greater detail in sequential compared to simultaneous lineups, a possibility that was examined in Experiments 1-3.

4) Does the evidentiary threshold for identification differ depending on whether the alternatives are presented together or one at a time?

If decision criterion placement is at a higher level in sequential compared to simultaneous lineups, then target faces that are positively identified from a sequential lineup should match the study face on a larger number of features on average than those that are positively identified from a simultaneous lineup. In addition, if discrimination is better in sequential lineups, then sequential participants might be better able to recognize a face using fewer features than simultaneous participants. In Experiments 4 and 5 we tested these possibilities.

THE RELATIONSHIP BETWEEN LINEUP MEMBER SIMILARITY AND ACCURACY IN SIMULTANEOUS AND SEQUENTIAL LINEUPS

Experiment 1

Introduction

In the current study, the effects of lineup similarity structure on identification accuracy were examined. Participants studied a list of faces and then attempted to identify them from a lineup, which was presented either simultaneously or sequentially. Showup tests were also utilized, and results from these were compared with the lineup identification outcomes to estimate the degree to which the foils enhanced or detracted from recognition accuracy. The lineups contained two dimensions of variation: similarity of the foils to the study face (foils-study), and similarity of the target to the study face (target-study). The logic behind the manipulations chosen is that a relative compared to an absolute judgment process should be more sensitive to variations in the similarity structure of the lineup.

For the first dimension manipulated in the lineups, targets were placed either in a lineup of foils that were matched to the study face on a particular facial feature or in a lineup of foils that were all randomly chosen from a large pool of faces. On average, the foils and the study face should be more similar in matched compared to random lineups following this method. We further surmised that the odds that at least one face in the lineup would match and/or be relatively more similar to the study face in memory should increase as the foil-study similarity increased. Therefore, the rate of choosing any face should be higher in the matched compared to random condition for both sequential and simultaneous lineups.

However, a change in the choice rate might be due to a change in foil or target identifications. It was expected that increasing the foil-study similarity should increase primarily foil choices in sequential lineups, while increasing foil-study similarity should increase primarily target choices in simultaneous lineups. The reasoning behind these predictions was as follows: If previously seen foils do not influence target identifications in sequential lineups, then the rate at which the target is chosen should not be affected by foil similarity. That is, the likelihood that the target surpasses the decision criterion should be equal in high and low similarity lineups. With respect to the foils, increasing their similarity to the target should increase the odds that one of them surpasses the decision criterion. Therefore, in sequential lineups, which are thought to involve an absolute recognition process, increasing the similarity of the foils should not impact target selections, but should impact foil selections. Consequently, in matched compared to random lineups, foil choices should increase and target selections should not change in the sequential condition. With respect to simultaneous lineups, if a relative decision process is in effect, the similarity of the lineup alternatives to the study face should influence the rate at which the target is chosen. If the foil-study similarity is comparably low, the target should be relatively more familiar and therefore chosen at a higher rate. Under low and high similarity conditions in simultaneous lineups, the target will be relatively the most familiar; therefore, foil choices should not be affected by lineup member similarity.

In addition to manipulating foil-study face similarity, the second dimension of the lineups varied was target-study face similarity. Regardless of lineup procedure employed, we of course expected that targets that were identical to the study face would be chosen at a higher rate than low similarity targets. A low similarity target, however, might emerge as the best candidate for selection if the foils are comparatively lower in their similarity to the study face. In this regard, a relative judgment strategy, which necessarily entails making comparisons among the faces, should increase low similarity target identifications over the rate found for judgments arrived at through an absolute strategy.

We were also interested in testing the possibility that features are analyzed differently depending on whether the test faces are presented together or one at a time. Specific facial features, such as the eyes, which are important in face recognition (Haig, 1984), might be analyzed more extensively in sequential compared to simultaneous lineups because the witness knows that the picture is going to be displayed only once. Moreover, false identifications of the feature substituted target should be reduced if the he no longer has that particular feature. To investigate, we varied within the foil matched lineups whether just the eyes, mouth, or face shape of the foils matched the study face.

Lastly, strength of admonishment against making false identifications was manipulated in the present study to test whether hits and false alarms would be differentially affected depending on identification procedure. If the primary difference between the two procedures is that sequential participants are withholding making a choice because their decision criterion is set at a higher value, then identifications of both the identical and lower similarity target should be less frequent in sequential compared to simultaneous presentations. Higher criterion placement in sequential compared to simultaneous presentations also leads to the prediction that admonishment manipulations in sequential lineups will have a larger effect on hits than false alarms, while for simultaneous lineups, admonishment manipulations should have a larger effect on false alarms compared to hits. Moreover, decreasing the similarity of the target to the study face should lower positive identifications of the target to a greater extent in simultaneous compared to sequential lineups when a higher level admonishment is introduced.

Method

Participants

A total of 294 people participated in exchange for course credit. One participant was removed from the final analysis due to experimenter error. The final sample was 86% female, and in total, 58% of the sample was Asian, 34% Caucasian, 6% Hispanic, and 2% self-identified as other.

Design

The identification procedure (simultaneous, sequential, and showup), admonishment level (low and high), and lineup feature similarity structure (random and matched) variables were fully crossed to create 12 conditions to which participants were randomly assigned. Each participant was given 12 study faces and accompanying identification tests. The similarity of the to-be-identified lineup face to the study face was controlled within participants. For half of the identification tests, the identical[1] study face was presented. For the other half, a highly similar looking face was presented instead. Lastly, we nested within the matched lineup similarity structure condition the facial feature used to match the foils in the lineup to the study face (eyes, mouth, or face).

Stimulus Materials

Face stimuli were created using FACES, a composite drawing software program used by law enforcement. Combined over all types of facial features (eyes, nose, mouth, face, eyebrows, jaw and head shape, hairstyle and facial hair), the program contains a databank of over 3850 facial components that can be used to construct a composite drawing. Within each facial feature category, the user can select from a range of morphological sizes (i.e., the default size of a chosen feature, such as the nose, can be enlarged) and types (i.e., the shape of a particular feature, such as bulging or almond shaped eyes). In addition, distance between features (e.g., distance between the two eyes, distance between the nose and mouth) can be controlled. The program can also generate at random a complete composite face. To create the study faces and accompanying lineups in the present study, a pool of 1000 faces was created by having

---

[1] In a pilot study, we presented participants with lineups in which the target face was a mirror image of the face that had been studied. Our purpose in so doing was to ensure that participants were making their identifications based on their recognition of facial features rather than on some other more general pattern matching strategy. However, we learned in post study interviews that participants were reluctant to make an identification if they did not find a face that precisely matched the study face. For instance, if a slight fold appeared under the right eye only, mirroring the face resulted in the fold appearing under the left eye. In such cases, we found that when participants recognized a face as a mirror image of the study face, many would not identify it because it was not a perfect match. In fact, a small number asked spontaneously during test whether they should identify a mirror image. Thus, in the final study reported here, test pictures were not mirrored from the original. In addition, the size of the lineup test pictures was reduced to about a third of the size of the study pictures to make identifications based on pattern matching alone more difficult.

the program generate faces at random. All distinguishing features, facial hair, and head hair were removed from each face.

Simultaneous and Sequential Lineups. The lineups for the random similarity structure condition were formed by randomly selecting without replacement 6 faces from the initial pool of 1000. The only restriction on the faces selected for the matched lineups was that the eyes had to be the same color (light or dark) across members. From these 6 faces, the to-be-identified study face was randomly designated. Using this procedure, 12 study faces and 12 accompanying lineups were created.

For the matched similarity structure condition, a similar procedure was employed to create the lineups, though rather than the features across the lineup faces varying at random, the lineup faces were matched on one feature. As before, six faces were randomly selected from the stimulus pool. After one of the six was designated at random as the target face, three versions of the lineup were created. Each version was based on the particular feature (eyes, mouth, and face shape) that the foils had in common with the target face (see Figure 2 for an example of the lineups used for one of the study faces). In each feature match condition (eyes, mouth, and face shape), the foils matched the target on only one feature; the remaining features for each of the foils were the same as the original version selected from the stimulus pool. In the eye matched condition, for instance, each of the foil faces in the lineup was altered from the original such that the eyes were now identical in shape and size to the target face. Note that in this condition the eyes were now identical across all foils in the lineup. Using this procedure, 12 study faces and three accompanying lineups for each were created. Participants in the matched condition were randomly assigned 12 lineups at test, with the constraint that they view one lineup for each study face and a total of four lineups for each of the 3 feature match conditions.

In both the random and matched conditions, the target appeared early in the array (positions 1-3) for half of the study faces and late (positions 4-6) for the other half. The order in which the foils appeared in the lineups was randomly determined. The position of the target and foils in each lineup was held constant across subjects. Furthermore, for the three versions of matched lineups made for each study face, position of the lineup members was maintained. If the lineup test was simultaneous, the pictures were presented in 2 rows of 3 pictures. The faces presented at study were about 18 x 16 cm and the individual pictures presented in the lineup at test were 5.5 x 6.5 cm in size. A number corresponding to the position of the face in the lineup (1-6) was placed beneath each picture for purposes of the identification task. For the simultaneous lineups, numbering was from left to right, starting on the top row.

To manipulate the similarity of the lineup target to the study face, a highly similar version of the study face was created for both the random and matched conditions. This new version of the study face was created by deleting one feature from the original study face and substituting it for another. The single feature deleted was the eyes, face shape, or mouth. The feature substituted in place was randomly determined. Therefore, a highly similar looking face that matched the study face on all features except one was made to create a target absent lineup condition for each study face (see Figure 2 for examples).

To create the target absent lineups used in the random similarity structure condition, study faces were randomly assigned to have one of the three types (eyes, face shape, and mouth) of feature deletions. Therefore, the study face and the feature used to create the accompanying target absent version of the lineup were confounded in the random condition.

In the matched condition, the feature deleted was counterbalanced across the 12 study faces. The altered feature was identical to the one that had been used to match the foils. For example, if the foils had the same eyes as the study face, then the suspect look-a-like would have different eyes than the study face (and hence the foils in the lineup).

For the random and matched lineups containing the highly similar target, the study face was removed from the lineup and the similar looking target was placed in the same position. Each participant was given six target absent lineups, and of these, two were of each feature substitution type. In half the lineups, the similar looking target appeared early and in the other half late.

Showups. For both the random and matched conditions, the faces designated as the targets for the simultaneous and sequential lineups were presented alone as a showup test. Half of the tests presented the identical study face, whereas the other half presented a similar looking target.

Measures

Overall accuracy was calculated separately for the identical and similar looking target conditions for the simultaneous (n=94), sequential (n=93), and showup (n=134) conditions. In the identical target condition, selecting the target was coded as an accurate response, while selecting a foil or rejecting the lineup was coded as an incorrect response. For the similar looking target condition, rejecting the lineup was coded as an accurate response, while selecting a foil or the similar looking target was counted as an incorrect response.

Three more dependent measures were calculated to further analyze responses in the simultaneous and sequential conditions, including: 1) Choice rate, or the mean proportion of responses in which any face was positively identified; 2) target choice/ID, or the mean proportion of responses in which the target face was chosen given the participant positively identified any face in the lineup; and 3) confidence level, or the average level of confidence reported. These dependent variables were calculated within participants for each target similarity level (identical and highly similar) and then between participants at each of the lineup similarity by admonishment conditions. Additionally, for participants in the matched lineup condition, choice rate and target choice/ID were calculated at each of the three feature match conditions (eyes, face shape, and mouth).

Procedure

At the start of the study, participants in all three identification test conditions (simultaneous, sequential, and showup) were simply asked to pay close attention to the faces that they were about to see, that they would be asked questions about them later. Study faces were then presented in random order for 10 s each followed by a 5-min retention interval, during which time participants completed a crossword puzzle. Thereafter, participants in the low admonishment condition were verbally instructed with the following:

> If you do not think that the study face is present, do not pick anyone. Please be aware that on any given test, the study face may or may not be present.

For those in the high admonishment condition, they were given the previous instruction plus the following additional information:

> Remember, as in a real life identification situation, the previously learned face, or “guilty” party, may or may not be present. It is important that you understand that you do not have to identify anyone. The fact that you are viewing a lineup does not necessarily mean that there is a previously studied face present. It is crucial that you do not identify a person unless you think you studied the face. An incorrect identification would be the equivalent of sending a truly innocent person to jail. In the case of a robbery, for instance, it is far worse to implicate an innocent person than to fail to identify a guilty person. Therefore, you must refrain from choosing a face unless you are absolutely certain that you previously studied the face.

Participants were correctly informed that the order in which the lineups (or showups) were presented did not necessarily match the order in which the faces had been studied. In addition, for the simultaneous and sequential conditions, they were told that there would be only one person, if any, in the lineup that was a previously studied face. In other words, there would never be more than one study face presented in a given lineup. All participants were asked to rate their confidence after making a response to a lineup (or showup) using a 5 point scale, with 1 being “not at all confident, just a guess” and 5 being “absolutely confident, would testify in court”.

Those in the sequential lineup condition were given additional instructions in keeping with prior research (see Wells et al. 1998). They were informed that for each lineup, they would only see one picture at a time. For each picture, they were to make a yes/no decision regarding whether the face had been presented at study. They were informed that once they made a decision, the decision would stand. They would not be allowed to view previously seen pictures, nor would be they be allowed to change any of their answers. In addition, if they positively identified a face, the test would terminate—even if there were faces in the lineup that remained to be seen—and the next identification test would be given.

For all identification conditions, testing was not timed. The experimenter waited until the participant made a selection before moving on to the next lineup. If a participant asked whether they made a correct decision, the experimenter would indicate that she did not know, or that she could not tell him until the end of the study.

All instructions were given verbally. Prior to running any participants through the protocol, the researchers running it rehearsed the instructions until they could be delivered without reading from the above script. Though experimenters were encouraged to use their own words and delivery style, they were instructed to communicate in a serious tone (just as they might imagine a law enforcement officer would in instructing a witness) and to cover all of the instructions in the above script.

Results

In this experiment and in all subsequent experiments, alpha was set at .05. In addition, as a measure of effect size, we report partial eta-squared.

Feature Substitution Effects in Showups

First we wondered if the feature substitutions made to the study faces would significantly decrease the familiarity of the target when presented alone in a showup. Participants were more apt to correctly reject the feature removed target (M=.72) than they were to correctly identify the identical target (.59), $F(1,132)=16.59$, $p<.01$, $\eta_p^2=.11$. This result also implies that showup participants were biased toward rejecting rather than positively identifying the test faces.

Study Face Memorability in Random and Matched Showups

Next we checked whether the targets utilized in the random compared to the matched condition affected the results. No significant effect of similarity condition on accuracy (random M=.64 versus match M=.66) in showups was obtained. Therefore, any differences between random and matched lineups that are reported in subsequent analyses cannot be attributed to pre-existing differences between the two conditions in the memorability of the study faces.

Accuracy in Simultaneous and Sequential Lineups

Adding foils to the lineup seemed to reduce correct responses, as mean accuracy was significantly lower in sequential (.41) and simultaneous lineups (.46) compared to showups (.65), $F(1,317)=82.49$, $p<.01$, $\eta_p^2=.21$. In showups, positive identifications of the study face were significantly higher, while identifications of the feature substituted target were significantly lower than in simultaneous or sequential lineups (all p's<.05). These results indicate that target discrimination was affected by the foils, regardless of whether they were presented simultaneously or sequentially.

As the means presented in Table 1 suggest, participants were less likely to reject a simultaneous compared to sequential lineup (M=.33 versus M=.43, respectively), $F(1,183)=6.85$, $p<.01$ $\eta_p^2=.04$. For those who did make an identification, they were more apt to identify the target if they viewed the lineup simultaneously rather than sequentially (M=.38 and M= .26, respectively), $F(1,183)=20.62$, $p<.01$, $\eta_p^2=.10$. In discovering why these differences emerged, we began our analysis by looking at how the structure of the lineup affected choice rates depending on identification procedure.

First we examined whether choice rates were influenced by the similarity of the foils. For both types of identification procedures, a higher rate of choosing occurred if all of the foil features varied rather than if one feature was matched across all foils (simultaneous: random M=.68 versus match M=.63; sequential: random M=.62 versus match M=.54). This was evident in the significant main effect obtained for lineup construction, $F(1,183)=4.91$, $p<.05$, $\eta_p^2=.03$. The direction of the effect was opposite to that predicted, however, as participants viewing random lineups were on average more likely to choose a face than those tested with matched lineups.

We turned to the accuracy data to determine if the average rate at which the target was chosen was influenced by the similarity of the surrounding foils to the study face. If participants are using a relative judgment strategy, suspect choices should increase if the foils are less similar to the study face. The overall rate of choosing the target (as opposed to rejecting the lineup or choosing a foil) was significantly affected in both identification conditions by the similarity of the foils to the study face when the similarity of the target to the study face was relatively low. Specifically, the feature substituted target was picked at a significantly higher rate if he was in a random compared to matched lineup in the simultaneous condition, $F(1,92)=6.65$, $p<.05$, $\eta_p^2=.07$, as well as in the sequential condition, $F(1,91)=9.79$, $p<.01$, $\eta_p^2=.10$. With respect to the identical target, participants were more likely to select him if he was in a random compared to matched lineup if a simultaneous procedure was used, though the effect was marginally significant, $F(1,92)=3.35$, $p=.07$, $\eta_p^2=.04$. The rate of selecting the identical target in sequential lineups was unaffected by similarity condition, $p=47$.

ID target/choice data were examined next to examine under what conditions targets drew more choices than foils. If a relative judgment strategy is operating, for those who picked a face, the rate of choosing the study face rather than a foil should be higher in random compared to matched lineups. In simultaneous lineups, the study face was not chosen significantly more often in random ($\underline{M}$=.71) compared to matched (.64) lineups, $\underline{p}$=.29. For people who made a choice from a sequential lineup, the rate of selecting the study face was also not affected by similarity structure of the lineup, $\underline{p}$=.51. In both random and matched lineups, sequential participants who made a choice selected the study face a little more than half of the time ($\underline{M}$=.57 in both similarity conditions). Examined next was the rate of choosing the target when the similarity of the target to the study face was decreased by feature substitution. The feature substituted target was not chosen significantly more often in random compared to matched lineups in neither the simultaneous condition nor the sequential condition, p's>.27.

Position Effects

According to the means presented in Table 1, simultaneous participants who chose from the matched lineups were more likely to choose the identical target rather than a foil, while sequential participants were just about as likely to pick a foil as they were the identical target. What led to the comparatively higher rate of target over foil identifications in simultaneous compared to sequential lineups? We thought that the position of the target in the lineup might account for the difference. Specifically, if the target was presented late within a matched sequential lineup, perhaps he was more likely to be missed than if he was presented earlier in the sequence. In other words, targets appearing late would be missed more often because a foil presented earlier in the sequence was chosen instead. We found that position did have a significant effect on choosing the target; the average rate of selecting the target was higher in earlier (M=.28) compared to later (M=.22) positions, $F(1,66)=4.23$, $p<.05$, $\eta_p^2=.06$. The rate of choosing any face from the lineup was identical in early and late positions (M=.53 for both positions). Among those who made a positive identification, on average 52% of the time the suspect was chosen in the early condition, while only 46% of the time the target was chosen in later positions. Furthermore, this difference does not seem to be due to some general tendency in sequential lineups to choose earlier rather than later in the sequence, because the rate of choosing in sequential random lineups was nearly identical in early compared to later positions (M=.30 versus M=.31, respectively), $p=.77$. In addition, the position effect found for sequential lineups does not seem to be due to the particular targets that were placed in early compared to later positions, as no significant position effects were obtained in simultaneous matched lineups (early M=.34 versus late M=.32), $p=.37$.

Admonishment and Accuracy Rates

Previous analyses demonstrated that participants were more likely to choose in simultaneous compared to sequential lineups, a finding that is consistent with the view that participants on average set a lower decision criterion in simultaneous compared to sequential lineups. Moreover, simultaneous participants were more likely to identify the target (rather than reject the lineup or identify a foil) compared to their sequential counterparts, $F(1,132)=11.06$, $p<.01$. This effect did not depend on target similarity to the study face, as target level did not significantly interact with lineup procedure, p=.82. In addition, an identification procedure effect was evident even when the data were analyzed at the level of study face, wherein accuracy rates were collapsed across participants to gauge how often on average a particular study face was chosen in a simultaneous versus sequential lineup. In 78% of the simultaneous lineups (n=36), the study face was identified more often than one of the foils (i.e., greater than or equal to 50% of the subject witnesses viewing the lineup selected the target). In sequential lineups, the study face was selected more often than a foil in 64% of the lineups. With respect to “innocent suspect” identifications, in 36% of the simultaneous lineups he was selected more often than a foil, and 28% of the time if presented sequentially.

Assuming that the lineups in which the target face was identical to the study face (target identical) were more familiar on average than the lineups that had the highly similar looking target (target similar), initial criterion placement should have different consequences for target identical and target similar lineups depending on admonishment instruction. In particular, if sequential participants are already biased to reject the lineups, then admonishment to use an even stricter criterion should decrease the rate of choosing in identical target lineups to a larger extent than in target similar lineups. This hypothesis received support, as the rate of choosing in sequential target identical lineups was significantly decreased by issuing a stronger admonishment (low admonishment M=.69 versus high admonishment M=.59), $F(1,91)=4.46$, $p<.05$, $\eta_p^2=.05$, while in sequential target similar lineups, choice rates were marginally affected (low admonishment M=.53 versus high admonishment M=.44), $F(1,91)=3.15$, $p=.08$,

η_p^2=.03. In simultaneous lineups, admonishment to use a stricter criterion should have the opposite effect. That is, since initial criterion placement is lower in simultaneous lineups, instructions to use a more conservative decision criterion should result in choice rates decreasing to a larger extent in target similar than in target identical lineups. This hypothesis received some support, as admonishment had a marginal effect on choosing in simultaneous target similar lineups (high admonishment M=.51 versus low admonishment M=.62), $F(1,92)=2.99$, $p=.08$, $\eta_p^2=.03$, and had no effect on choosing in simultaneous target identical lineups (high admonishment M=.70 versus low admonishment M=.75), $F(1,92)=1.46$, $p=.23$, $\eta_p^2=.02$.

Effects of Feature Matching on Identification

Choice rate data for lineups (collapsed across identification procedure) are displayed in Figure 3. Results are presented first for the participant level analysis. A target level analysis will follow.

First we asked whether the particular feature removed from the test face in the absence of any foils influenced accuracy within participants. A one way ANOVA on the choice data in showups suggested that the particular feature removed did affect choosing, $F(1,218)=3.44$, $p<.05$, $\eta_p^2=.03$. The mouth (M=.28) and eye (M=.30) removed targets were selected at a higher rate than the face shape altered (M=.20) targets. The face shape, therefore, might be more important in face recognition than the other features, as if a study face is altered by changing the shape of the face then recognition is more likely to fail. Would placing the target among foils that were matched to the study face on a particular feature alter this effect?

Consistent with the showup data, the feature manipulation did have a similar effect on choice rates in lineups. There was a marginal difference in the rate of choosing any face from the lineup if the face shape of the target had been substituted compared to when the eyes or mouth had been substituted, $\underline{F}(1,134)=3.49$, $\underline{p}=.06$, $\eta_p^2=.02$. This pattern was consistent across the lineup identification procedures.

Next we analyzed whether accuracy in participants who chose depended on the feature used to match the foils. The showup results led to the prediction that the face shape substituted target should be selected less often than the foils who match the face shape of the study face. This analysis also enabled us to establish whether we had any evidence that simultaneous and sequential participants weigh features differently during recognition. That is, the results should diverge between the two features if they are analyzing the faces differently. A total of 20 participants from the simultaneous and 23 from the sequential condition chose at least once from each of the target level by feature conditions (i.e., there were two trials of each target X feature combination and the participant chose from at least one of them across all target X feature combinations). The target choice/ID scores of these participants were entered into a four-factor (2 target level X 2 admonishment level X 3 feature condition X 2 identification procedure) mixed design analysis of variance to determine whether the rate of choosing the target differed depending on identification conditions. The only factor that significantly affected target choice/ID scores was the similarity of the target to the suspect, $\underline{F}(1,39)=10.59$, $\underline{p}<.01$. If the target was identical to the study picture, he was chosen more often compared to when one of his features was substituted for another ($\underline{M}=.51$

versus M=.33, respectively). Feature condition and lineup procedure were not significant, nor did they interact with any variables.

To allow us to examine within study faces how foil feature manipulations influenced recognition, we analyzed the choice and target choice/ID data over subjects. As shown in Table 2, identification rates varied considerably across the 12 study faces. In addition, no statistically significant feature effects were found on choice or target choice/ID rates at the level of study face. The showup data predict that foils should be selected more often than the target in the face shape condition compared to the other two feature conditions. As shown in Table 2, foils were picked more often than the feature substituted target in 7/12 targets in the face condition, 8/12 targets in the mouth condition, and 10/12 targets in the eye condition. The target level analysis, therefore, appears inconsistent with prediction. Across targets the standard deviation for the means obtained in each feature condition were more variable across targets in the face compared to the eye and mouth conditions, suggesting that the extent to which feature manipulations have an effect on identifying a face from a lineup might interact with other factors, such as how distinctive the faces are in the lineup.

Confidence

Confidence scores were entered into a mixed design ANOVA, with identification procedure and target as the independent factors. Confidence was higher when the identical rather than feature substituted target was present in the lineup, $F(1,132)=5.41$, $p<.05$, $\eta_p^2=.04$. In addition, sequential participants were more confident on average in their identifications than simultaneous participants, $F(1,132)=11.19$, $p<.01$, $\eta_p^2=.08$.

Identification procedure did not interact with target level, p=.45. Accuracy and confidence were correlated within each identification condition, and it was found that when the identical target was present, the relationship was stronger in sequential (r=.43, p<.01, two-tailed) compared to simultaneous (r=.18, p=.08, two-tailed) lineups. The reverse was true when the feature substituted target was present: Confidence and correct rejections were significantly related in simultaneous lineups (r=.30, p<.01), but not in sequential lineups (r=.02).

Discussion

In the current experiment, we were interested in determining whether the similarity structure of the lineup had a different effect on the distribution of choices depending on lineup procedure. The relative judgment model predicts an increase in target identifications as the foils are decreased in similarity to the perpetrator. In such cases, the target will be chosen because he is relatively more similar to the perpetrator than the other faces. If an absolute judgment strategy is operative, however, then the structure of the lineup should have little consequence for the rate at which the target is identified.

Overall, adding foils to create a lineup decreased discrimination, as accuracy was greater in showups compared to simultaneous and sequential lineups. These results suggest that target discrimination is decreased by adding foils, regardless of whether they are presented simultaneously or sequentially. In addition, we found evidence that participants in both identification conditions were influenced to some extent by the similarity structure of the lineup. Specifically, decreasing the similarity of the foils to the study face significantly increased positive identifications of the feature removed target in simultaneous and sequential lineups.

If participants were simply comparing the face presented to the representation of the study face in memory, foil similarity manipulations should not influence the rate at which the target is chosen. Our results indicate that false alarms can be induced in sequential lineups by simply selecting fillers that are low in similarity to the perpetrator. This could result in a contrast effect of sorts, wherein the feature substituted target face is segregated with greater ease from the background of foils. The foils, which were randomly chosen for the lineup, should have on average a lower familiarity value compared to the target than the foils in the matched lineup. In addition, that the effect occurred in sequential lineups suggests that the lineup faces are compared in memory during sequential presentations.

If participants were computing figure/ground differences, then the study face should have been chosen at a higher rate than the feature substituted target in a random lineup. Specifically, the difference between the matched foils and the target is smaller than the difference between the random foils and the target. Contrary to this prediction, when the study face was present in the lineup, the rate of selecting him was not increased by having dissimilar foils in the sequential administrations, and only marginally increased target selections in simultaneous administrations. There is at least one possible explanation for why the study face was not chosen at a higher rate in random compared to matched lineups. Since participants were presented with many study faces, it is conceivable that some of the foils in the random lineups were similar to and therefore mistaken for a different study face than the one for which the lineup was created. If interference of this sort were occurring, it should have had a greater effect in sequential lineups. That is, sequential participants that might have otherwise identified the study face do not because a foil that seemed similar to another one of the study faces was identified first. In simultaneous lineups, such errors are minimized by displaying all of the pictures at once.

With respect to the finding that participants were not more likely to select faces from matched compared to random lineups, the results were not as expected. Increasing the similarity of the foils to the study face did not significantly draw choices away from the target, nor did it increase the rate at which participants chose any face from the lineup. When the foils were relatively high in similarity to the study face, participants in both identification conditions tended to reject the lineup rather than pick a foil, though to a greater extent in sequential lineups. Perhaps this had to do with the fact that the foils in the lineup all had a particular feature in common, and hence had the same familiarity value. If the target himself was not familiar enough to be chosen by the participant, then the odds that one of the foils would exceed the threshold for identification would on average not only be lower than the odds that the target would exceed the threshold, but also equivalent across the foils in the lineup.

Finally, sequential participants were more biased toward rejecting the lineups than simultaneous participants. Moreover, in keeping with the signal detection analysis, admonishment level seemed to have a larger effect on false alarms for simultaneous procedures and a larger effect on hits for sequential lineups. The results also suggested that participants examine the features of the faces in a similar manner when they are presented simultaneously and sequentially. The particular feature used to match the foil to the study face did not have a different effect on accuracy depending on lineup procedure. Additionally, the particular feature substituted on the study face did not alter the rate at which he was identified from the array in both procedures.

Experiment 2

Introduction

This experiment employed the removal without replacement procedure to test whether simultaneous compared to sequential participants are more likely to select the relatively most familiar foil when the target face is removed from the lineup.

The random and matched lineups from the first experiment were utilized so that the distribution of target and foil choices from the identification tests (hereafter referred to as "full lineups") in Experiment 1 could be compared to the distribution of choices made in Experiment 2 when the target was removed (hereafter referred to as "removed lineups"). The pattern of choices in the full and target removed lineups was compared to similarity rankings of the foils compared to the target made by another group of participants. Previous studies on the target-foil shift in target removed lineups have shown an increase in selections of the foil most often identified from the full lineup (Wells, 1993; Clark & Davey, 2005). However, the most widely identified foil might not necessarily be the most similar foil with respect to the perpetrator. For instance, if a foil was particularly more distinctive than the others it might be subjectively viewed as more familiar and hence identified more often than the others. In this case, the foil is not selected based on memory but for other reasons. Thus, an independent measure of foil-culprit similarity is in order.

In both lineup procedures, if the most highly ranked foil corresponds to the most often identified foil in the full lineups, then this suggests that participants in both procedures are cognitively processing the lineup faces in a similar manner. In other words, if participants in sequential lineups utilize different information than their simultaneous counterparts, then the rate of agreement on foil identifications and the most highly ranked foil should not be correlated between the two procedures. Such regularity would not be expected if different match to memory information was used during recognition in sequential compared to simultaneous lineups.

Lastly, we had the raters make their rankings either based on their memory for the target or with the target picture in view. A recent meta-analysis of eyewitness identification studies found that when experimenters use similarity ratings to determine whether a lineup is fair, they always have participants make their ratings of the foils with the study face in view (Flowe, Ebbesen, Burke, & Chivabundtt, 2001). This procedure is valid to the extent that the face information being compared visually is similar to the face information compared in memory during the identification task. It is conceivable that when a lineup face is compared to a perpetrator in memory, different face information is used compared to when a side-by-side or perceptual comparison is made with both the study and foil face in view.

Method

Participants

A total of 344 undergraduates participated for course credit. Among these, 72 participated in the identification portion of the study, while 272 rank ordered the faces in the lineups.

Design

The protocol for the recognition part of the study was similar to Experiment 1, with identification procedure (simultaneous and sequential), admonishment (low and high), and lineup similarity (random and matched) fully crossed to form eight experimental conditions to which participants were randomly assigned.

With respect to the similarity ranking portion of the study, another group of participants were provided with a target face and told to rank order the members in the corresponding lineup from most to least similar. Whether the rank orders were given based on their memory for the target or while they were looking at the target during the task was controlled between participants. Ranking condition (memory and perception) was crossed with lineup similarity structure (random and matched) to create four conditions to which participants were randomly assigned. Within the matched condition, participants compared the study face to the lineup foils half of the time, while the other half they compared the feature substituted target to the lineup foils. In the random condition, participants always compared the study face (never the feature substituted target) to the foils in the lineup. In addition, participants in the matched condition were randomly assigned to make rankings for 12 lineups, with the constraint that they view 1 lineup for each study face and a total of four lineups for each of the three feature match conditions. Those in the random condition viewed the 12 random lineups that were presented in Experiment 1. In both the random and matched conditions, the lineups were presented to each participant in a random order.

Recall that for each study face in the matched condition, there were three lineup versions, which varied in the feature that the foils had in common. The foils across the three lineup versions for a given study face were identical on all features except on one of the features (eyes, mouth, and face shape) that was being varied in each lineup. For each study face, a minimum of 40 participants ranked each version of the corresponding lineup (half saw the identical target while making the rankings, half the feature substituted target). In the random condition, 24 participants ranked each lineup.

Procedure

The protocol used in the identification portion of the study was similar to that employed in the first experiment. Participants were told to pay attention to the study faces presented, as they would be asked questions about them later. They studied a total of 12 faces for 10 s each, and then after a 5-minute retention interval, they were admonished (high or low) and tested on target removed lineups presented either simultaneously or sequentially.

In the ranking portion of the study, participants were asked to rank order the lineup members from 1 to 5, with 1 designated as “most similar” and 5 “least similar” to the target face. Tied rankings were not allowed. Those in the perceptual condition were given the target face and were allowed to look at it as often as they wished while completing the ranking task. Those providing the rankings based on their memory for the target, viewed the target faces for 10 s each one after the other. After a 1-min retention interval, the target removed lineups were presented one at a time and participants gave their rankings. A total of 14 participants were removed from the analysis because instructions were not followed (e.g., assigned tied ranks, did not rank all of the faces).

Results

Comparisons of Ranking Conditions and Ranking Across Features

To determine which lineup foil was the most similar to the target face, the ranking data within each lineup (random n=12, matched n=72) were averaged across raters. The lineup member with the highest average rank was deemed the most similar to the target. Following this procedure, a total of 60 lineups had corresponding memory ranks and 79 had perceptual ranks, as some lineups had to be excluded because foil choices were on average ranked equally. In addition, the average ranking for the most highly ranked foil was used as a measure of how consistently the foil was ranked at the top across participants. For instance, an average rank of "1" indicates that the most highly ranked foil was ranked by every participant as the foil most like the target. As the average rank increases, this indicates greater variability across participants in which of the foils was the highest ranked.

Next we compared the memory rankings with the perception rankings. The most highly ranked foil in the memory ranking condition was the most highly ranked foil in the perception condition in 52% of the lineups. In addition, the average ranks for the most highly ranked foil in the memory and perception condition were positively correlated ($\underline{r}$=.22, $\underline{p}$<.05). A mean comparison of the average ranks found that the ranks given in the perception condition were more variable than those given in the memory condition, with the average perception ranking being smaller than the average memory ranking (1.91 versus 2.09), $\underline{F}(1,83)=8.92$, $\eta_p^2=.10$. This suggests that the criteria used to make the rankings might have been more variable in the perception compared to the memory condition.

We also examined in both the matched and random lineup conditions whether the ranking method used produced different results. The average perception and memory rankings did not differ in the random lineups, whereas they did differ in the matched lineups, $\underline{F}(1,69)=11.72$, $\underline{p}<.01$, $\eta_p^2=.14$. In matched lineups, the average ranking was larger in the perception ($\underline{M}$=2.12) compared to memory ($\underline{M}$ =1.90) condition. In addition, within the matched condition, the particular feature used to construct the lineup differentially influenced how variable the rankings were, $\underline{F}(2,69)=5.16$, $\underline{p}<.01$. The average rankings (across the ranking methods) seemed to be more variable in the face ($\underline{M}$=2.17) compared to the eye ($\underline{M}$=1.89) or mouth ($\underline{M}$=1.98) condition. Followup analysis found that only the face and eye conditions significantly differed from each other ($\underline{p}$<.05, Tukey's HSD test). It seems then that whether the face shape was taken into consideration in assessing similarity varied across subjects to a larger extent than consideration of the eyes. Lastly, feature did not interact with the particular ranking method used to determine the most similar foil, $\underline{p}$=.48.

Ranking Data Compared to Foil Identifications

Next we checked how often the similarity rankings corresponded to foil choices in Experiment 1. The foil identification data from Experiment 1 were first coded at the level of lineup to determine which of the foils from each was most often chosen. Next we compared the ranking data with the most popular foil choices by lineup for Experiment 1. Lineups that had 2 or more foils tied for the highest identification rate and/or tied for the highest ranking were removed from the analysis. For the full lineup identification data, 19 simultaneous lineups were removed and 11 sequential lineups were removed. For the ranking data, 5 memory ranked lineups were removed and 4 perceptually ranked lineups were removed. After removing the necessary lineups, we found that the most popular foil when ranked from memory was the most often identified foil 34% of the time (15/43) in simultaneous lineups and 31% of the time (20/64) in sequential lineups; these rates of agreement were not significantly different, $\chi^2(1)=.44$, $p=.50$. When the foil faces were ranked perceptually, with the target face in view, there was less agreement between the identification outcome and the rankings, and even more so in simultaneous compared to sequential lineups. In the perception condition, the most popular foil identified was also the foil receiving the highest ranking in 8% of the simultaneous lineups (4/48), and in 18% of the sequential lineups (12/64); these rates approached but were not statistically significant, $\chi^2(1)=2.43$, $p=.12$. The correlation between the two procedures in agreement between the foil rankings and identifications was positive and significant in the memory condition ($r=.56$, $p<.01$, two-tailed), and almost significant in the perception condition ($r=.29$, $p=.07$, two-tailed).

We also performed the analysis with the lineups that had tied rankings and foil identifications and the pattern of results was the same. The coding scheme was altered in the following manner to perform this analysis: If one of the foils among those receiving the highest rank was the same as the foil (or one of the foils) that was most often identified, the lineup was coded as an agreement case. Some lineups, however, still had to be excluded from the analysis because no foil was a more popular choice than the others (i.e., the foils that were identified were identified at the same rate). After using this more liberal coding scheme, we found for the memory ranks that the identification data agreed 46% of the time (31/67) in simultaneous lineups, and 35% of the time (27/78) in sequential lineups, and the rates did not differ based on procedure $\chi^2(1)=2.03$, $p=.15$. As for the perception ranks, the agreement rate was 27% for simultaneous lineups (19/67) and 23% for sequential lineups (18/79), again not a significant difference in the rates for the procedures, $\chi^2(1)=.27$, $p=.59$. The correlation between the two procedures in agreement between the foil rankings and identifications was again positive and significant in the memory condition ($r=.35$, $p<.01$, two-tailed), but not in the perception condition ($r=.12$, $p=.35$, two-tailed). Taken together, these results suggest that when people choose from a simultaneous lineup, the choices tend to be spread over a greater number of members, and hence the "most popular foil" designation goes to more than one foil more often in simultaneous than sequential lineups. Lastly, these data also illustrate that memory compared to perceptual similarity rankings more closely correspond to identification outcomes, perhaps even more so in sequential lineups.

Choices in Removed Lineups by Lineup Procedure

Next we analyzed the removed lineups to determine the degree to which the most popular foil in Experiment 1 was chosen at a higher rate once the target was removed. The data were analyzed at the level of study face (collapsed across the 3 feature conditions) to increase the probability that one of the foils would emerge as the clear favorite. Based on this coding, the most popular foil in simultaneous lineups was the most popular foil in sequential lineups for 50% of the study faces in the matched condition (see Table 3), and for 25% of the study faces in the random condition (see Table 4).

The rate at which the most popular foil was chosen in the full lineup and the removed lineup were treated as a within subjects factor and entered into a mixed ANOVA, with lineup similarity and identification procedure as the between subjects factors. The most popular foil was chosen at a significantly higher rate in the removed compared to the full lineup, $F(1,36)=10.62, p<.01, \eta_p^2=.22$. Lineup procedure did not have significant effect overall on the rate of choosing the popular foil, though an interaction between lineup procedure and target removal approached significance, $F(1,36)=3.20$, $p=.08$, $\eta_p^2=.08$. Inspection of the means indicated that the popular foil was chosen at a lower rate when the target was present in simultaneous compared to sequential lineups ($M=.14$ versus $M=.18$, respectively); however, when the target was removed, selection of the most popular foil was nearly the same in the two lineup procedures ($M=.22$ versus $M=.20$, respectively). These results suggest that when the target is removed, simultaneous participants direct their choice from the target to the most similar foil.

Additional analyses on the rate at which the remaining foils (i.e., those besides the most popular foil) were chosen when the target was removed suggests that participants in both lineup conditions, but especially in simultaneous lineups, are more likely to pick foils when the target is removed. The rate at which other foils were chosen in the full lineup and the removed lineup were treated as a within subjects factor and entered into a mixed ANOVA, with lineup similarity and identification procedure as the between subjects factors. A significant interaction was obtained between target removal and identification procedure, with the rate at which other foils were chosen increasing in both lineup conditions when the target was removed, but more so for simultaneous participants, $\underline{F}(1,44)=16.37$, $\underline{p}<.01$, $\eta_p^2=.27$. Specifically, other foils were chosen at an average rate of .15 (SE=.02) when the target was present and .34 (SE=.03) when the target was removed in sequential lineups, and .18 (SE=.02) when the target was present and .50 (SE=.03) when the target was removed in simultaneous lineups.

Lastly, lineup rejection data were also entered into a mixed ANOVA with identification procedure and similarity structure as between groups factors. The only significant effect was lineup procedure, $\underline{F}(1,44)=16.19$, $\underline{p}<.01$, $\eta_p^2=.27$; the main effect for lineup construction approached significance, $\underline{F}(1,44)=3.12$, $\underline{p}=.08$, $\eta_p^2=.07$. Participants were more likely to reject the lineup if it was sequentially rather than simultaneously presented ($\underline{M}=.45$ versus $\underline{M}=.34$, respectively), and rejections were also independently higher for matched compared to random lineups, (M=.42 versus M=37, respectively).

Discussion

The purpose of the current study was two fold: First, to determine whether similarity rankings could predict identification outcomes for both simultaneous and sequential lineups, and second to determine whether the most often identified foil in the full lineups corresponded to the most identified foil in the removed lineups regardless of identification procedure. We found modest levels of agreement between the similarity rankings and identification outcomes for both procedures. Specifically, the most often identified foil tended to be the highest ranked foil relative to the target in both simultaneous and sequential lineups. This suggests that participants in both lineup conditions use similar memory to match evidence in making an identification.

Comparison of the target removed identifications with the full lineup identifications found that simultaneous and sequential participants selected the most popular foil at equal rates when the target was removed. Additional analysis of other foil choices in the lineup found higher rates of other foil selections when the target was removed, especially in simultaneous lineups. The difference between the two lineup procedures in the rate of selecting other foils is attributable to the greater tendency to reject the lineup if a sequential procedure is used.

Consistent with Experiment 1, the rejection rate was higher when a matched rather than random lineup was used. If they determine that each face seems to match the study face to the same degree, and on average none more so than the others, then the lineup is rejected. Further support for this notion comes from the fact that there was less agreement among people ranking the lineups in the matched compared to the random condition. Intuitively this outcome makes sense because in the matched lineup, the foils are matched to each other and the study face on one feature; therefore, the foils in the lineup should each be about the same in similarity to the target. Thus, there should be greater variability in which one of them is ranked as the top foil among people viewing the lineup.

Interestingly, the ranking data suggest that perhaps the proper metric used to determine how similar the members of a lineup are should involve memory based similarity ratings of some kind. The current experiment found that the memory rankings were a better predictor of identification outcomes than those involving a direct visual comparison of the target and lineup members. This suggests that the features that are retained in memory and/or privileged during the recognition task might differ from those that are analyzed when similarity ratings are based on a direct visual comparison of the faces. For instance, across the visual comparisons there might be a high degree of variability in the particular features that are used to judge the similarity of the members compared to when memory based comparisons are made.

Experiment 3

Introduction

The next study sought to explore whether participants compare lineup faces in memory while viewing sequential lineups. In so doing, the mock witness testing paradigm (Doob & Kirshenbaum, 1973) was employed using the face stimuli from the first experiment. The mock witness test entails presenting participants with lineups without having them previously study any faces. Participants are asked to select any face from the lineup that seems to stand apart from the others. The mock witness test is used to detect whether the lineup is biased against the target/perpetrator. The logic is that if participants are able to pick out the target at greater than chance levels without out having studied him, then the lineup is biased against the target either for structural reasons (e.g., the target picture is slightly tilted away from the other pictures) or because the lineup foils do not match the target on a major feature (e.g., the target is male and the foils are female), or the foils are in some way highly dissimilar to the target.

For our purposes, first we wanted to know whether choice rates in the absence of memory for the study face would be lower in sequential compared to simultaneous lineups. That is, after obviating the absolute judgment process by using the mock witness procedure, would the differences in choice rates observed in the first experiment be eliminated? Second, we wanted to determine whether lineup similarity structure affected mock witness choices differentially depending on identification procedure. In Experiment 1, participants in both lineup conditions chose the feature substituted target at a higher rate if he was placed in a random compared to matched lineup. Since sequential participants seem to be influenced by the similarity structure of the lineup in the same manner as simultaneous participants, this opens up the possibility that participants in the sequential task are comparing in memory the previously seen lineup faces. We predicted in the mock witness task that the highly similar target should be chosen more often in the matched compared to random lineup. In the matched lineup, the foils all have the same feature in common, whereas the feature substituted target does not have that feature in common with the surrounding foils. Hence, the feature substituted target looks less similar on the whole compared to the others, and should be chosen more often in a matched compared to random lineups, in which none of the features across the target and foils are matched. Furthermore, if sequential mock witnesses select the target from the matched lineups at a higher rate than from random lineups, this suggests that while making a yes/no decision regarding a particular face, information that was extracted from previous faces is being compared in memory.

An absolute judgment strategy should also reduce the likelihood that witnesses are affected by how fair the lineup is to the target, as judged by the quality and number of foils in the lineup. Two types of lineup measures have been developed to determine how fair a lineup is. These measures are based on the distribution of choices across the lineup obtained from the mock witnesses (see Malpass & Lindsay, 1999 for a review). One type of measure, lineup size, attempts to determine the number of "good" foils there are in the lineup. The fewer the number of foils there is, the greater the likelihood that the suspect can be chosen at random. The other type of measure, lineup bias, attempts to discern the degree to which the lineup biases witnesses toward picking the suspect rather than another lineup measure. We calculated these measures for our lineups, and predicted that if participants tend to adopt an absolute recognition strategy in sequential lineups, then the lineup size and bias measures should not be related to the pattern of identifications made in Experiment 1. That is, sequential participants should be just as likely to choose a suspect from a lineup regardless of whether the number of acceptable lineup members is small or large, or whether the lineup is biased toward or away from the suspect. Simultaneous participants, if they are using a relative judgment strategy of sorts, should be affected to a greater extent by these factors.

Since we also presented the mock witness task sequentially, we wondered whether the lineup size and bias measures obtained with sequential mock witnesses might postdict accuracy in sequential lineups. Past research has compared identification rates found for sequential lineups with lineup fairness measures that were based on simultaneous rather than sequential mock lineups. Lindsay, Smith, and Pryke (1999), for instance, performed such an analysis and found that measures of lineup size did not predict false identifications in neither simultaneous nor sequential lineups. Furthermore, lineup bias measures predicted false identifications in only simultaneous not sequential lineups, leading them to conclude "lineup measures appear to postdict false identifications only for inferior lineup procedures" (p. S101).

There are a number of ways in which lineup size and bias measures taken on the same lineup might differ depending on whether the mock witness task is simultaneous or sequential. First, intuitively it would seem that sequential mock participants might identify faces that are presented earlier rather than later in the sequence. Hence, the lineup size measures—which involve in their calculation an examination of the rate at which each member of the lineup is chosen—should be appreciably smaller in the sequential compared to simultaneous mock witness procedure because sequential participants are less likely to evaluate the last members in the lineup sequence. Second, with regard to using lineup size to postdict false identifications, perhaps only when the target is presented early in the lineup would lineup size be predictive. If the innocent suspect is presented later in the lineup, he might be missed because an earlier presented foil is chosen instead. Consequently, the correlation between the rate of false identifications of the innocent suspect and lineup bias measures will necessarily be on the low side. Thus, we predicted that lineup bias measures might postdict accuracy when the innocent suspect is presented earlier rather than later in the lineup.

Finally, investigators who assess the fairness of the lineups they use in their identification experiments typically use either mock witness ratings or pairwise similarity ratings (Flowe et al., 2001). In the current experiment, we had another group of witnesses rate the similarity of the members of the lineups using multiple rating methods, including ratings made to faces presented 2 at time, or all 6 faces at once, or to all 6 faces after viewing them sequentially. Our reason in so doing was to determine which rating method best predicted suspect choices, and whether the particular method that was the most predictive depended on lineup procedure.

Method

Participants

A total of 312 undergraduates participated in exchange for course credit, including 132 in the mock witness portion, 120 who gave pairwise similarity ratings, and another 60 who made global similarity ratings.

Design

For the mock witness task, participants were randomly assigned to one of four conditions, which were created by crossing identification procedure (simultaneous and sequential) with lineup feature similarity structure (random and matched). Within the matched conditions, half of the lineups contained targets that matched the foils on one feature, whereas the other half contained targets that did not match the foils on any features. Within the random condition, half of the participants received lineups in which the study face was present, and the other half lineups in which the feature substituted target was present.

For the similarity ratings, another group of participants was randomly assigned either to give a single rating to all six faces in the lineup taken together (global similarity) or to give ratings to all possible pairs of faces from the lineup (pair-wise similarity). Each participant rated 12 lineups, half of which were target present. Within each rating condition, half of the participants rated the random lineups, while the other half rated the matched lineups. Moreover, in the matched condition, they evaluated four of each feature condition type (eye, mouth, or face matched foils).

Half of the global similarity participants evaluated the matched lineups sequentially while the other half evaluated them simultaneously (the minimum number of participants rating each lineup in each identification presentation condition was 16). In the pairwise condition, across the 72 matched lineups (12 study faces x 3 feature conditions x 2 target levels), a total of 1080 pairs (15 possible pairings x 72 lineups) were rated by a minimum of 15 participants each. Across the 12 random target present lineups, a total of 180 pairs (15 x 12) were rated by 20 participants. Thus, the pairwise similarity rating task produced over 19,800 data points.

Similarity Measures

The lineup size and bias measures were calculated for each study face using the mock witness data. The measures were calculated by conditioning the mock witness data on identification procedure (simultaneous versus sequential) and target level (identical versus feature substituted target) in both the matched condition (collapsed across the 3 feature levels) and in the random condition. Lineup size was measured using two standard methods of assessment, including the effective size calculation (Malpass & Lindsay, 1999) and by using the number of acceptable lineup members technique (Malpass & Devine, 1983). Effective size basically involves adjusting the nominal size of the lineup downward if there are foils that do not receive an adequate number of choices by mock witnesses (see Malpass & Lindsay, 1999 for details). The number of acceptable lineup members is calculated by counting the number of foils in the lineup that are chosen by mock witnesses at a rate that is at least 75% of chance expectation. Additionally, three commonly used measures of lineup bias were taken, including functional size (Wells, Leippe, & Ostrom, 1979), defendant bias (Malpass, 1981), and whether the target was selected at above chance expectation (Doob & Kirshenbaum, 1973). Functional size is a measure that indicates the extent to which a lineup is biased toward or away from the defendant/study face, and it is calculated by simply taking the reciprocal of the proportion of witnesses who chose the defendant/study face. Defendant bias is a test of whether or not the number of mock witnesses choosing the study face exceeds the number of expected by chance, with the level of chance expectation based on the effective size (i.e., number of foils drawing mock witness choices) rather than the nominal size of the lineup. The last bias calculation is another binary test of whether the target is chosen above chance expectation, except that the level of chance expectation is based on the nominal size of the lineup (i.e., the proportion of witnesses selecting a given foil in a lineup at chance expectation is .17 since there are 6 members in the lineup).

As for the pairwise similarity data, four measures were calculated for each lineup for each participant making the rating, including the average target-foil similarity rating, the average foil-foil similarity rating, the difference between the average target-foil similarity and average foil-foil similarity ratings, and lastly the standard deviation for the pairwise ratings. These four similarity measures obtained for the pairwise ratings were then averaged across all participants for each target's respective lineup. With respect to the global similarity ratings, the ratings of individual subjects were averaged together for each target conditioned on whether the ratings were obtained after viewing the faces simultaneously or sequentially. Additionally, all similarity ratings, whether pairwise or global, were conditioned on whether the study face or feature substituted target was in the lineup.

Procedure

In the mock witness portion of the study, participants were told that they would be evaluating lineups that were previously used in a face recognition study. The experimenter explained that another group of participants had studied a set of faces and then had their recognition memory for the faces tested with lineups. In the present study, their task was to try to determine which face, if any, was the target face in each of the lineups. They were instructed to not identify anyone if none of the faces seemed to stand apart from the others, because in the previous study, sometimes the study face was not present in the lineup.

In the sequential condition, participants were further told that they would view the lineup faces one at a time and make a yes/no decision for each face that they saw. Once they identified a face from a lineup, no more pictures from that lineup would be shown. They were also told that once they gave a response, their response would stand. In other words, they were not allowed to change an answer once it was given.

For the participants in the pairwise similarity condition, all possible pairings of the members in each of 12 lineups was presented. In total 15 similarity comparisons were made per lineup. Across the 12 lineups evaluated, therefore, each participant rated the similarity of 180 pairs of faces (15 comparisons x 12 lineups). The 180 pairs were presented in random order, one pair at a time on the computer screen, with the faces in the pair present simultaneously throughout the rating task. Ratings were made by placing the mouse pointer on a similarity scale (that ranged from 0 to 100, with 0 being "not at all similar" and 100 being "completely the same") on the part of the scale that corresponded to their desired rating. The rating appeared in an onscreen textbox, and the rating could be changed as needed by re-clicking the scale. When satisfied, the rater would click a button to submit the rating. The experimenter walked participants though three practice trials, encouraging them to base their ratings on the facial features of the lineup members. Thereafter, participants made the ratings on their own and at their own pace.

Participants in the global similarity condition provided their rating of how similar the faces were by naming a number from 0-100, with 0 being "not at all similar" and 100 being "completely similar". In addition, a picture of a lineup containing highly dissimilar members (i.e., the spatial arrangement of the features differed drastically across members) was placed under the low end of the scale, and a picture of a lineup containing all clones was below the high end of the scale. Participants had this scale on hand throughout the rating task. The lineups were presented by the experimenter either simultaneously or sequentially. Raters made the ratings individually and at their own pace.

Results

Mock Witness Data

Figure 4 illustrates the outcome of the mock witness study. First we asked whether lineup construction had a different effect on choice rates overall, depending on whether the mock witness test was given simultaneously or sequentially. Choice data were entered into a two-factor (2 identification procedure X 2 lineup structure) between subjects analysis of variance. The choice rate was higher on average if the lineup was presented simultaneously (M=.83) rather than sequentially (M=.75), $F(1,128)=4.03$, $p<.05$, $\eta_p^2=.02$. Participants were also more likely to choose if given a matched (M=.83) rather than random lineup (M=.74), $F(1,132)=5.31$, $p<.05$, $\eta_p^2=.02$. Since no interaction was obtained between identification procedure and lineup construction (p=.50), it seems that increasing the similarity of the members in a lineup increases willingness to choose in both types of lineup procedures. Thus, in both procedures witnesses seemed to use the similarity of the faces to gauge the likelihood that the target was in the lineup.

Next we wondered if the rate at which the study face was identified by mock witnesses was predicted by identification procedure and lineup member similarity. The rate of choosing the suspect in the random lineup was not significantly larger than the rate expected by chance ($\underline{M}$=.167), $\underline{t}(62)=1.17$, $\underline{p}$=.13. The rate of selecting the suspect from a matched lineup, however, did differ from chance, $\underline{t}(67)=-3.36$, $\underline{p}<.05$, suggesting that the similar looking foils in these matched lineups drew choices away from the suspect. Moreover, identification procedure did not significantly predict alone or in conjunction with lineup structure the rate of identifying the target.

The rate of target identifications in the matched lineup condition as a function of target similarity and lineup procedure was examined next using a two-factor (2 target level X 2 identification procedure) mixed design analysis of variance. If participants are comparing faces, then the target should be chosen at a higher rate if he is less rather than more similar to the foils. Participants did select the feature substituted target who did not match the foils on any features at a higher rate ($\underline{M}$=.17) than the target who matched the foils on one feature ($\underline{M}$=.13), though the effect was not significant, $\underline{F}(1,64)=2.53$, $\underline{p}$=.12. In addition, the direction and size of this effect was the same for both types of identification procedures, suggesting that participants were attending to figure/ground differences during both identification procedures.

Lineup Fairness Measures

We turned next to the lineup size and lineup bias measures to discover whether the rate at which participants picked the target during a lineup recognition test could be postdicted. First we examined whether the lineup size and bias measures differed depending on the mock witness identification procedure used (simultaneous or sequential). Lineup size measures did indeed vary depending on whether mock witnesses were tested with the pictures all at once or one at a time. The average number of acceptable lineup members, $F(1,44)=9.39$, $p<.01$, $\eta_p^2=.18$, and the average effective size, $F(1,44)=22.19$, $p<.01$, $\eta_p^2=.33$, was smaller in the sequential compared to simultaneous condition, indicating that choices made by sequential participants were spread over fewer lineup members compared to simultaneous participants. In 9/12 of the random lineups, participants selected faces early in the sequence.

Only 1 of the 3 lineup bias measures could be analyzed because only one target (while in a sequential lineup) in all of the lineup procedures and target levels was identified by mock witnesses at a level beyond chance expectation (calculated based on lineup nominal size), and no targets were identified at a level above chance using the defendant bias criteria. With respect to functional size, the remaining lineup bias measure that could be analyzed, no differences emerged due to lineup condition; however, differences were found depending on target level and lineup similarity structure. The functional size data indicated that participants overall were more biased toward picking the target if he was in a random compared to matched lineup (functional size random $M=6.73$ and functional size matched $M=8.86$), $F(1,32)=4.14$, $p<.05$,

η_p^2=.11. In addition, lineup similarity significantly interacted with target level, $F(1,32)=8.99$, $p<.01$, η_p^2=.22. Inspection of the functional size means suggests that in matched lineups, substituting a feature on the target that did not match the foils tended to bias participants away from picking the target (identical target M=7.00, substituted target M=10.72). In random lineups, feature substitutions made to the target biased participants towards picking the feature removed target (identical target M=8.00, feature substituted target M=5.47).

The lineup size and bias measures, which were calculated separately for sequential and simultaneous mock witness data, were correlated with positive identifications of the target from Experiment 1. The results are shown in Table 5. As can be seen, none of the measures significantly predicted positive identifications of the target when in a simultaneous or sequential lineup. The distribution of mock witness responses also seemed to differ based on whether the procedure was administered simultaneously or sequentially, as none of the lineup size and bias measures were significantly correlated across the mock witness identification conditions.

Overall, the distribution of mock witness choices seemed to have little post predictive value in determining the rate at which the targets were identified during the lineup recognition test. Would the pairwise or global similarity ratings fare any better in this regard? The pairwise and global similarity measures were correlated with choice and target choice/ID rates, conditioning the analysis on identification procedure and target level (n=12 target present random lineups, n=36 target present matched lineups, and n=36 target absent matched lineups). In Experiment 1, the study face was identified just as often from a matched compared to a random lineup. When a feature was removed from the study face, however, the target was identified more often in a random lineup. Therefore, a low correlation between accuracy and similarity was expected when the target was present, while a negative correlation was expected with the target was absent.

The pattern of correlations obtained is in keeping with the prediction. Table 6 provides the correlations obtained between accuracy and each of the similarity measures by target level and lineup procedure. Accuracy was measured in two ways: the rate at which the suspect was identified overall, and the rate at which the suspect was identified for choosers. The pattern of findings was the same for both accuracy measures. When the study face was present, none of the similarity measures predicted the identification of the suspect in either of the identification procedures. When the target was absent, however, the similarity of the foils to each other, and the target to the foils predicted target choices in both procedures, indicting that participants in the simultaneous and sequential conditions were responding to the similarity structure of the lineups in much the same way.

For simplicity, S-F difference scores were computed by subtracting the average target-foil similarity ratings from the foil-foil similarity ratings. If the difference is positive, this suggests that the target was more similar to the foils than the foils were to each other, while a negative value indicates that the foils were more similar to one another than the target was to the foils. The fact that S-F ratings were significantly lower for feature removed (M=-4.75) compared to identical targets (M=-0.15) supports this idea, $p<.02$. If the correlation between accuracy and the S-F difference is positive, this suggests that having highly similar foils in a lineup deters witnesses from picking the target. In contrast, if the suspect is similar to the foils and the foils are not as similar to one another, the suspect will be picked out at a higher rate. For both procedures, the difference scores were positively correlated with accuracy overall (all p's$<.05$).

Overall, the correlations obtained for simultaneous lineups were higher than for sequential lineups, suggesting that the relationship between similarity and identification accuracy might be stronger if the pictures are viewed all at once. In addition, the correlations were stronger in target absent compared to target present lineups. Lastly, the lower the variability of the ratings within a lineup the more likely it was that the target was identified. Thus, it seems that the similarity ratings made to the members of a lineup predicted accuracy to a greater extent if the faces in the lineup were viewed as highly similar by the raters.

Interestingly, the global similarity comparisons that were made sequentially (M=45.93) did not significantly differ from those given while viewing all of the faces (M=46.04), p=.91. This result implies that more stringent criteria were not being used to evaluate a lineup presented sequentially compared to simultaneously. In addition, the similarity of the lineups were evaluated in a consistent manner across the two procedures, as the global similarity ratings were positively correlated (r=.37, $p<.05$).

Full Lineup Identification Rates by Study Face and Lineup Procedure

Finally, if the manner in which the faces are presented at test influences the number and/or type of features that are used in comparing the foils to the study face in memory, little agreement is expected between the two procedures in the particular study faces that are consistently recognized. That is, if the type of memory evidence involved in the recognition process differs, then faces that are better recognized from one identification procedure should not necessarily be better recognized in the other, all other things being equal (e.g., the lineup is fair, demand characteristics are equal across conditions). The average rate at which the targets were identified was positively correlated between the two identification procedures, $r=.61$, $p<.01$. This held true for targets in the matched condition ($r=.74$, $p<.01$) and in the random condition ($r=.61$, $p<.01$). Since the faces identified from simultaneous lineup tended to be the faces that were recognized in sequential lineups as well, this suggests that the evidence used to compare the faces in the lineup to the study faces in memory might be similar across the two procedures.

We further examined the target choice/ID data for each of the 12 study faces used in the matched condition to determine whether some were more often identified than others depending on lineup procedure. In the showup condition, 10/12 were remembered by more than 50% of participants. The two remembered below this rate were also not remembered by more than 50% of the witnesses in the two lineup conditions. In simultaneous lineups, 7/12 study faces were identified by more than 50% of choosers in every feature condition, while in sequential lineups, 6/12 were consistently recognized. Of the 6 consistently remembered study faces in the sequential condition, 5 of them were among the 7 consistently recognized in the simultaneous condition. With respect to the feature altered targets, sequential participants picked foils more often than the feature substituted target. When the lineup was sequentially administered, 6/12 feature altered targets were consistently never identified by more than 50% of participants in any of the 3 feature conditions. In other words, foils were consistently preferred in 6/12 targets. In contrast, only 2/12 substituted targets were never identified consistently from simultaneous lineups, and in fact 6/12 targets were identified by a majority of participants in every feature condition.

Therefore, it appears that in the time frames used in the present study, the presence of foils reduced the number of study faces that were consistently recognized. In addition, neither the particular feature used to match the foils to the study face nor the method used to present the lineup affected identification of about half of the study faces. Lastly, altering the familiarity of the target resulted in more foil identifications in sequential but not simultaneous lineups.

Discussion

In this study, our objective was to determine whether sequential participants compare the lineup faces in memory during the recognition task. If they do, then this opens the possibility that the similarity structure of the lineup will affect identifications made by sequential participants. We found that mock witnesses viewing lineups sequentially were influenced by the similarity structure of the lineup. Further analyses measuring the fairness and similarity of the lineup revealed consistent patterns between the two procedures in the way in which identifications of the target were influenced.

In the mock witness task, participants in the sequential condition were less likely to choose than those in the simultaneous condition. It seems that even in the absence of trying to identify a face in memory from a set a faces, choice rates are reduced when faces are evaluated sequentially. This might be the case because sequential participants are faced with the uncertainty of whether an even more distinctive choice is yet to come, and hence are reluctant to make a pick.

Mock witness choice rates were higher in matched compared to random lineups, regardless of lineup procedure employed. Contrary to the first experiment, choice rates might have been higher in matched lineups because it was the only cue available to mock witnesses to determine the likelihood that the study face was present. Since the rate of choosing was higher in matched lineups for both identification procedures, this opens the possibility that sequential participants are comparing the faces in memory.

In addition, the similarity structure of the lineup and the memorability of the study face seem to independently influence face recognition accuracy in both simultaneous and sequential lineups. Accuracy was correlated with similarity in both procedures. If an absolute judgment process is strictly operating, then the similarity of the lineup members should not influence accuracy. The results we obtained suggest that recognition is influenced by the lineup context in both procedures. Furthermore, it appears that the target is more likely to be identified if he is more similar to the foils than the foils are to each other. This relationship was stronger in if the target was feature removed. Under these circumstances, the foil-foil similarity was higher relative to target-foil similarity.

Lastly, it is interesting that recognition of the study face is not benefited by the similarity of the foils like recognition of the feature substituted target is. Perhaps this is because the recognition process occurs in two stages: First, a lineup face is compared to study face information recalled from memory. If the match exceeds a certain threshold for making a positive identification, then the face is positively identified. When the lineup target is identical to the study face, the face is more often than not identified at this stage. If the face does not exceed the threshold, then the face is reconsidered in relation to the other faces in the lineup. If it is more familiar than the others, it is identified. If not, then the lineup is rejected.

COMPARISON OF IDENTIFICATION THRESHHOLDS IN SEQUENTIAL AND SIMULTANEOUS LINEUPS

Experiment 4

Introduction

The previous studies demonstrated that sequential witnesses can be influenced by the similarity structure of the lineup in the same manner as simultaneous participants. In both procedures, placing foils in the lineup that are relatively low in similarity to the study face increases identifications of the feature substituted target, or "innocent" target. However, the rate at which the study face, or "guilty" target, was identified did not seem to differ depending on the similarity structure of the lineup. In addition, simultaneous participants were more likely to identify both "guilty" and "innocent" targets compared to their sequential counterparts, a finding that is in keeping with the signal detection analysis of accuracy in simultaneous and sequential lineups.

The purpose of the next two studies is to determine whether sequential participants require more feature evidence in making a positive identification from a lineup. We learned from previous studies that accuracy did not depend on the particular feature that was used to match the foils in the lineup to the study face, nor were differences in accuracy explained by simultaneous and sequential participants differentially considering particular features. In the next two studies, we consider whether the extent to which all features combined, not just any particular feature, exceed a decision threshold affects identification from lineups. We predicted that the threshold for making an identification would be relatively higher in sequential compared to simultaneous lineups. If so, this could explain why there is less of a difference between the two procedures in the rate at which the study face was identified, and more of a difference in the rate at which the feature substituted target was identified.

To test these ideas, we created the multiple lineup procedure, which entails presenting participants with a series of five lineups to test their recognition of a previously studied face (see Figure 5 for an illustration). As participants progress through the lineup series, the number of features that differ across the faces in the lineup increase. The goal of the participant is to correctly identify the study face as early in the series of lineups (i.e. with as few features varying across members) as possible. We predicted that overall participants would be more likely to choose and be more accurate if the faces in the lineup were presented simultaneously. In addition, we hypothesized that a greater number of features would be required in the sequential compared to the simultaneous lineup condition.

Method

Participants

A total of 100 undergraduates participated to fulfill a course requirement.

Design

A 2 (lineup procedure) x 2 (admonishment level) x 2 (memory strength) x 3 (initial lineup feature varied) mixed design was employed. The lineup procedure (simultaneous and sequential) and admonishment level (low and high) variables were fully crossed and controlled between participants. This resulted in four conditions to which participants were randomly assigned. Memory strength was varied based on duration of exposure time to the study face (10 s as low, 20 s as high).[1] Participants attempted to identify 6 study faces with the multiple lineup procedure; initial lineup feature varied (eyes, mouth, or face) was a within participants manipulation and counterbalanced across the 6 study faces for each participant.

Materials

Face stimuli were created using FACES. To create a lineup set in the present study, a random face was generated and all distinguishing features, facial hair, and head hair were removed. The resulting face served as a base face for creating the lineup. This base face also served as the clone participants viewed in the first lineup presentation. In the second lineup presented, eyebrows and noses were varied across all six faces, while all other features remained constant. For each successive lineup, one additional feature (eyes, mouth, and face) was varied, with the final lineup consisting of faces where all features varied across the six faces. The order in which features were introduced was counterbalanced across the 6 study trials for each participant. The position of the study face within the lineup was counterbalanced for each participant such that the perpetrator appeared early in the lineup half of the time.

A pool of 18 study faces and accompanying lineup sets (five for each study face) were created. In all, we created a total 864 faces, which were portrayed in 180 lineups. The particular set of faces that a given participant studied was determined by randomly assigning participants to one of five possible stimulus conditions. Each stimulus condition was created by randomly selecting six study faces and six accompanying lineup sets from the original pool of 18. Within the six lineup sets, two of each feature condition (i.e. the initial feature varied in the lineup was the eyes, mouth, or face) was given. Moreover, within each feature condition, the second feature introduced in the lineup series was counterbalanced across the two lineups (i.e. if mouth was the first feature introduced, then the second lineup in the series introduced the face next, while for the other mouth lineup set, the second feature introduced was the eyes).

A program was created in Visual Basic to display the study faces and lineup tests. In the sequential condition, the program displayed one face at a time and two buttons (one for "yes" and one for "no"). A practice session, with the appropriate lineup procedure instruction was created, to acquaint participants with the program. The onscreen position of the faces within the lineup was maintained in the sequential and simultaneous conditions to control for possible differences in accuracy that might arise from examining faces on different screen locations.

Procedure

Participants were seated approximately 40 cm away from the computer screen. A practice trial was run while verbally instructing participants individually. After each study face was presented, a "brainteaser" problem was given to create a retention interval of 1 min between study and test. Participants were told that the brainteaser problem was being given to help them pass the time, and that their answers would not be graded. Participants in the sequential condition were told to indicate for each face, whether or not it was the face that they had studied, while participants in the simultaneous condition were given instructions to either reject the lineup or identify a face by position number.

All participants were made aware that from one lineup to the next, the faces would become increasingly different, however the position of the faces within the lineup would remain the same. Once a participant made a decision about a particular face, or lineup, they could not change their mind, or review previous lineups, or faces. If participants identified a study face, the program would automatically move to the next study face. If the participant failed to identify a face, they were given another lineup, and they would keep seeing lineups until they identified the study face, or all five lineups had been exhausted. Subjects were told that it was better to identify the study face from the second lineup rather than the fifth, however accuracy was of equal importance.

Once the participant expressed an understanding of how the program worked, they were admonished (low or high) with the same instructions used in the previous experiments.

Measures

Rate of choosing any face (choice rate) across the six lineup sets was computed. If the participant chose at any step within the series, they were coded as having chosen from the lineup. Accuracy was calculated by determining for each participant the average proportion correct for each feature condition for the trials in which the participant made a choice. We also computed for each feature condition for which the participant made a choice the average number of features (2, 3, 4, or 5) that varied across the faces in the lineup to form a measure of feature variability.

Results

Data were analyzed with a mixed model ANOVA for the three major dependent variables: choice, target choice/ID, and evidence level. Alpha was set at .05 for all analyses.

Since we did not counterbalance feature condition or the order in which particular features were introduced across the study faces, first we checked whether choice or accuracy rates varied across the five stimulus sets to which participants were randomly assigned. We found no main effect of stimulus set on choosing or accuracy; however, a stimulus set x feature condition interaction was obtained for both dependent measures, indicating that the effects of the feature manipulations were inconsistent across the study faces. Therefore, the feature variable was not included in subsequent analyses, as the effect that it had seemed to be highly dependent on peculiarities of the to-be-identified face.

Choice Rates

Similar to the previous studies, participants in the simultaneous condition were more likely to choose a face than those in the sequential condition ((M=.72 and M=.43, respectively), $F(1,92) = 25.11$, $p<.01$, $\eta_p^2=.21$. Comparison of these choice rates with those obtained in Experiment 1 suggests that the multiple lineup procedure may have induced sequential participants to adopt an even stricter criterion. In Experiment 1, sequential choice rates were on average .64 in matched lineups and .63 in random lineups. Choosing in simultaneous lineups was not affected by the procedures used in the current study, as the choice rates were nearly identical between the two experiments (Experiment 1: matched M=.71 and random M=.75; current experiment M=.72).

The strength of memory manipulation only influenced choosing in simultaneous, not sequential, lineups, $F(1,92) = 15.10$, $p<.001$, $\eta_p^2=.14$. Longer exposure to the study face affected choosing on average in simultaneous (.55 for low versus .89 for high), but not in sequential (.48 for low versus .38 for high) lineups. Thus, it seems as longer exposures to the study face did not persuade sequential participants to choose at a higher rate.

Target Choice/ID Accuracy

A total of 37 simultaneous participants and 17 sequential participants chose at least once in every feature condition. The data for these participants was analyzed to determine if there were any differences in the rate at which the target was identified. Target Choice/ID accuracy by lineup condition and admonishment level is shown in Figure 6. In keeping with the findings obtained for full lineups Experiment 1, participants in the simultaneous condition were significantly more accurate (50% of the time) than participants in the sequential (34% of the time) $F(1,46) = 4.88$, $p<.05$. These rates are comparably lower than those obtained in the first experiment, wherein the average target choice/ID rate was .71 in random and .64 in matched simultaneous lineups, and .57 in both random and matched sequential lineups.

A marginally significant interaction was obtained between admonishment and lineup procedure $F(1,46) = 2.707$, $p<.15$. This interaction occurred because participants given the simultaneous lineup presentation were more accurate in the high admonishment condition (M=.59 in high and M=.41 for low) whereas participants in the sequential condition did not seem to be as greatly affected by admonishment instruction, (M=.31 in high versus M=.37 low). Perhaps this effect arose because the multiple lineup procedure caused sequential participants to set an even higher decision criterion compared to the standard lineup procedure; thus, there was not much room for the criterion to increase following the high admonishment.

Strength of memory did not significantly influence accuracy for choosers, though the effect was in the expected direction (high M=.47, and low M=.37).

Choice Level within the Multiple Lineup Series

The findings so far replicate the first experiment in which simultaneous participants were more accurate. Next we asked whether sequential participants chose later in the series than simultaneous participants. If so, this would mean that sequential participants required more features to establishing identity, and even with more features, they were still less accurate than simultaneous participants.

Contrary to expectation, participants in the simultaneous condition chose significantly later in the lineup series than those in the sequential condition, $F(1,46) = 8.32$, $p<.01$, $\eta_p^2=.15$. We further examined the data by conditioning accuracy on the level at with participants chose by lineup procedure. Our purpose in so doing was to determine whether simultaneous participants were more accurate overall because they tended to progress further in the series, and hence had a greater number of features than sequential participants on which to base their identification decision. The results indicated that mean accuracy was higher in simultaneous compared to sequential lineups across all feature levels (see Figure 7), suggesting that the higher accuracy rate in simultaneous lineups was not due to the fact that simultaneous participants tended to see more features, but rather because of the way in which they processed the available features.

Lastly, admonishment also had a differential effect on where participants chose in the series depending on lineup procedure employed, $F(1, 46) = 5.18$, $p<.05$, $\eta_p^2=.10$. Sequential lineup participants tended to chose sooner under high compared to low admonishment (low M=3.4, high M=2.9) while simultaneous participants chose later under high compared to low admonishment (low M=3.5, high M=3.9). Choosing later in the sequence was expected under the high admonishment instruction, as participants are told to not make an identification unless they are certain. Intuitively, certainty should increase as available information to make the decision increases. Maybe the sequential participants who made a choice in the high admonishment condition were more confident that they could identify the study face, and they thereby chose earlier in the sequence relative to sequential participants in the low admonishment condition. This is supported by the fact that admonishment affected whether the lineup was rejected in the sequential condition, but not the simultaneous condition.

Discussion

In the current experiment we set out to discover whether sequential participants required more evidence before choosing a face from a lineup. We found just the opposite: Of the participants who chose a face, simultaneous participants progressed further in the lineup sequence than sequential participants before making a choice. This finding suggests that perhaps simultaneous participants needed a larger number of features to compare and rule out faces. Hence simultaneous participants progressed further in the series, and hence tended to be more accurate. But, after controlling for the feature level at which they chose, participants were still more accurate in simultaneous lineups; therefore, it appears that with fewer features, being able to compare them enhances recognition.

There are other possible explanations, however, though for why recognition was higher in the simultaneous compared to sequential conditions. First, perhaps accuracy was lower in sequential procedure because they tended to choose prematurely. One reason why they might have is because sequential participants had to work harder than simultaneous participants if they wanted to view all of the faces in a lineup. Perhaps our subjects were not willing to work hard, hence they chose sooner in the sequential procedure so as to avoid having to make so many decisions. Choosing sooner in turn led to a lower accuracy rate. However, the fact that sequential participants were significantly less likely to choose any face poses a problem for this interpretation of the findings. That is, sequential participants were more likely to view every face in the lineup and see multiple lineups than they were to positively identify a face. A second explanation related to the first that is viable, however, is that the experimental procedures made it easier for simultaneous compared to sequential participants to correct possible identification errors. Suppose that a simultaneous participant was certain enough to choose a face early on in the lineup sequence, but chose not to, pressing further in the sequence to verify the face that had been tentatively chosen. In so doing, the face that they had chosen is viewed with yet another feature added. If they indeed had made an error, they would have the opportunity now to correct their decision and pick a different face. Sequential participants with a choice in mind, however, could not verify it with the same amount of ease as a simultaneous participant. Consequently, simultaneous participants could more easily correct for recognition errors than sequential participants, and therefore were more accurate overall.

In the current study we allowed participants to self-determine the level of information necessary before an identification was made. This enabled us to see the recognition process unfold in the two lineup procedures. However, the fact that the accuracy and feature level rates were correlated with the particular procedure used to administer the lineup made it difficult to conclude with certainty whether greater evidence is required in one lineup procedure compared to the other.

Experiment 5

Introduction

This experiment controlled systematically the number of features that participants could use in making an identification to disentangle the effects of individual differences and lineup procedure from the level at which participants made a choice. If sequential participants require more evidence before making an identification than simultaneous participants, then they should be less likely to choose any face when few features vary across faces. We also manipulated whether participants were forced to choose a face from the lineup. If we found under forced choice conditions that sequential participants can identify the study face at the same rate as simultaneous participants, this would mean that sequential participants deliberately miss faces that they recognize because they are seeking a face that is even closer in comparison with the face that they have in memory. Thus, this would suggest that the underlying recognition process is the same between the two lineup procedures, but that sequential participants miss targets they recognize because they are withholding choice.

Method

Participants

A total of 56 participants participated for course credit.

Design

Feature level, or the number of features that varied across lineup members (2, 3, 4, or 5), was controlled within participants and choice level (forced to choose somebody, or told to not choose anyone unless they were absolutely certain) was controlled between participants.

Materials and Procedure

Participants were to study four faces and make four identification attempts from target present lineups, one of each feature level type. The study faces used in the current study were randomly selected from the materials used in Experiment 4. From the original pool of 18 study faces, we chose a random sample of 16 to serve as study faces in the present study. Four stimulus conditions were formed by randomly dividing the 16 chosen faces into four groups of four faces each. Participants were randomly assigned to one of the four stimulus conditions. In order to control feature level with subjects, it was necessary to have one lineup for each feature level within each stimulus condition. We therefore randomly determined how many features (2, 3, 4, or 5) would vary within the lineup that corresponded to a particular study face, with the constraint that 1 study face in each stimulus condition would represent each feature level. Consequently, within each stimulus group, feature level was confounded with study face (though because we had four stimulus conditions, this helped us to increase generality).

Participants studied the four faces for 10 s each, and after a 5-min retention interval, were administered four lineups either simultaneously or sequentially. Participants in the forced choice condition were told that the study face was present and that the task required that they pick a face out of the lineup. Those in the no choice condition were told that though the study face was present, but that it was better to reject the lineup than to pick out the wrong face.

All participants were given practice lineups so as to demonstrate how the features would vary across lineup members. For instance, if only three features varied, the remaining features across the faces were held constant (i.e., the other features were identical). They were told to ignore any features that were identical across the faces, concentrate on the ones that differed, and to identify the one face that had features identical to one of the faces that they had studied.

Results

First we determined whether the particular stimulus condition to which participants had been assigned influenced choice or accuracy rates. We found that stimulus condition interacted with feature level (p=.02), but did not interact with any of the other independent variables (p's>.43). The stimulus condition by feature level interaction indicates that the extent to which feature level affected identifications depended on the peculiarities of the particular faces in the lineups. Because stimulus condition did not interact with lineup procedure or choice level, we collapsed across the four stimulus conditions in the analyses that follow.

In the accuracy analysis that follows, a lineup response was coded as incorrect if the participant did not make a choice.

First we analyzed the effect that the choice level had on choosing in the two identification conditions (see Figure 8). Participants in the simultaneous condition were significantly less likely make a choice if they were given the option not to choose ($\underline{M}$=.76) compared to when they were forced to choose ($\underline{M}$=.96), $\underline{F}(1,26)=13.00$, $\underline{p}<.01$, $\eta_p^2=.33$. In the sequential condition, choice level also affected the rate at which participants chose a face (no choice option $\underline{M}$=.69 versus forced choice $\underline{M}$=..86), $\underline{F}(1,18)=4.58$, $\underline{p}<.05$, $\eta_p^2=.15$. These findings also demonstrate that participants did not always choose a face in the forced choice condition, especially if they viewed the faces sequentially. Sequential participants were less likely to choose in the forced condition compared to simultaneous participants, though the effect was marginally significant, $\underline{F}(1,26)=3.66$, $\underline{p}=.06$, $\eta_p^2=.12$.

We examined within each identification condition the effect that both the feature and admonishment manipulation had on accuracy. Figure 9 illustrates the main results. In the simultaneous condition, a linear trend in accuracy based on the number of features that varied was found, $\underline{F}(1,26)=18.08$, $\underline{p}<.01$, $\eta_p^2=.41$. In addition, choice instruction did not affect accuracy, $\underline{p}=.72$, indicating that accuracy was not increased in simultaneous lineups by forcing participants to make a choice. With respect to sequential lineups, a linear trend in accuracy based on feature level was observed, $\underline{F}(1,26)=27.51$, $\underline{p}<.01$, $\eta_p^2=.51$. As predicted, sequential participants were more accurate if they were forced to choose (no choice option M=.30 versus forced choice option M=.45), $F(1,26)=3.56$, $p=.07$, $\eta_p^2=.12$.

Finally, accuracy in sequential and simultaneous lineups was compared in the no choice and in forced choice conditions. Under forced choice conditions, no differences in accuracy between the identification procedures were obtained (simultaneous M=.46, sequential M=.50), $p=.67$. If a no choice option was provided, the findings were in the expected direction (simultaneous M=.43, sequential M=.28), but they were not statistically significant, $p=.17$.

Discussion

This experiment was run to disentangle the effects of identification procedure and feature level from accuracy at were found in Experiment 4. After controlling the number of features that participants could use in making the ID, no differences in accuracy were obtained between the simultaneous and sequential lineups. Furthermore, if participants in sequential lineups were forced to choose a face, accuracy was significantly increased compared to the forced choice condition. In simultaneous lineups, accuracy was not affected by choice instruction. These findings suggest that simultaneous and sequential procedures do not significantly alter the way in which test faces are compared to the study face/perpetrator in memory.

GENERAL DISCUSSION

The current project endeavored to determine whether face discrimination differs depending on whether a lineup is presented simultaneously or sequentially. Prior research suggests that eyewitnesses viewing simultaneous lineups are likely to pick a face if it is relatively more familiar than the other faces in the lineup (Wells et al., 1998). In contrast, the alternatives are thought to not influence accuracy in sequential lineups because judgments are made based solely on whether a face is a sufficient match to the perpetrator in memory. In the series of experiments we reported, the effects of lineup member similarity on accuracy were evaluated to determine whether the foils influence face recognition in simultaneous compared to sequential lineups.

First, we found that presenting the target among distractor faces rather than alone significantly reduced accuracy. Table 7 presents the showup and lineup results obtained in Experiment 1, along with the results from a recently reported meta-analysis of showups (Steblay, Dysart, Fulero, & Lindsay, 2003) and lineups (Steblay et al., 2001). As shown, having distractors in the lineup decreased recognition accuracy, as hits were higher and false alarms to the feature substituted target were lower in showups compared to lineups. This finding is in keeping with a recently conducted meta-analysis comparing showups to lineups (Steblay et al., 2003). If witnesses were making strictly absolute judgments in sequential lineups, accuracy should have been more similar between showups and sequential lineups. According to our results, however, showups produced greater accuracy than sequential lineups.

We further hypothesized that the similarity of the distractors should influence accuracy if judgments are made based on a relative decision process. Evidence was found for relative responding in both sequential and simultaneous lineups. Contrary to the absolute judgment model, sequential participants were more likely to choose the feature substituted (i.e., "innocent") target if the foils in the surround were low in similarity to the study face. In addition, despite the fact that in the matched condition the feature substituted target stood apart from a background of foils that were uniformly similar, simultaneous participants did not seem to take advantage of the lineup structure. Mock witness identifications of the innocent suspect were not higher in simultaneous compared to sequential lineups. Taken together, these findings are in keeping with predictions that might be drawn from the WITNESS model (Clark, 2003), wherein familiarity is determined in part by the amount of info that can be extracted from a given face and matched to memory relative to other faces in the lineup. If the familiarity for a particular face is relatively high because the similarity of the next best alternative with respect to the perpetrator in memory is lower, then the face is chosen because subjectively it has a higher familiarity value.

The results also suggest that information is extracted from faces in a similar manner in both identification procedures. First, similarity rankings of the foils to the study face predicted which of the foils would be chosen most often in both lineup conditions. Second, consistent with Clark and Davey (2005), we found that the most highly ranked foil was chosen at an equal rate in simultaneous and sequential lineups. Third, faces that were better recognized in the sequential condition were also the ones that were better recognized in the simultaneous condition. Last, with respect to the feature manipulations of the foils and the target, the results across the two identification procedures were similar, suggesting that features affect the recognition process in a similar manner.

Several hypothesis tests were performed to determine if sequential participants are better able to discriminate the target from the foils. If target discrimination is better in sequential lineups, then recognition of the study face on fewer features than in simultaneous lineups should be possible. We systematically increased the number of features that could be used to make an identification, and found that simultaneous participants were slightly more accurate with fewer available features to discriminate the target from the foils than sequential participants. However, if participants were told that the study face was in the lineup and they were forced to make a pick, differences in accuracy between the two identification procedures were eliminated. This finding suggests that participants in both conditions are equally likely to pick the target out; accuracy varies between the two conditions primarily because sequential participants reject the lineup more often than simultaneous participants. Consequently, positive identifications of the suspect are lower.

In all of the experiments reported here, we found that lineup rejections were higher in sequential compared to simultaneous lineups. Sequential participants were less likely to identify any face from a full lineup (Experiments 1, 2, and 5), from a target removed lineup (Experiment 2), or during a mock witness task (Experiment 3). The mock witness results are especially important, as they show that even when the contents of memory are not searched for the study face, sequential participants still withhold picking a face. Witnesses might be reluctant to choose a face from a sequential lineup because they wonder if a better face will be presented later in the series. This result is in keeping with the signal detection analysis of identifications in simultaneous and sequential lineups put forth by Ebbesen and Flowe (2001).

Finally, the experiments reported extend previous work on lineup member similarity to sequential lineups. We found like previous studies that decreasing the similarity of the foils increases the rate at which the "culprit" is identified from a simultaneous lineup (Lindsay, Martin, & Webber, 1994; Wells, Rydell, & Seelau, 1993). No such increase was found in sequential lineups. In addition, if the foils were low in similarity to the perpetrator, the rate at which the innocent suspect is identified was increased in both identification procedures. This result is in keeping with prior research that has found that false alarms are increased by matching the foils to a description of the perpetrator (Clark & Tunnicliff, 2001; Lindsay, Martin, & Webber, 1994). These findings suggest that high similarity lineups should be used in order to protect innocent suspects. Position effects, however, create a problem for sequential administrations when the target is present. Specifically, like Clark and Davey (2005), we found that an alternative that is similar to the culprit might be chosen first if a sequential lineup is used and the target appear late in the series.

Though the present study utilized composite drawings as stimuli, the distribution of choices was remarkably similar to those obtained in face recognition studies (see Table 7). The only apparent difference is that false identification rates for the innocent suspect were higher in our study compared to previous studies. This is attributable to the fact that the innocent suspect that we used was high in similarity to the study face. Assuming that the findings would have been similar had actual faces been used, the variability in the identification rate across faces obtained in the present study underscores the importance of using multiple exemplars of culprits and foils in identification experiments. Identification outcomes and the subsequent conclusions that are drawn might be highly variable across studies otherwise.

The results of the current project are limited to lineups in which the foils are evenly matched to the perpetrator because the degree to which the foils are uniformly similar to each other has been found to influence identifications. Laugherty. Jensen, and Wogalter (1988) constructed lineup faces using composite drawing software, creating foil faces that matched the target face, but each of them on a different feature. As such, the foils were more different relative to each other than the suspect was compared to the foils. Participants then attempted to identify from these lineups faces they had studied previously. The lineups were actually blank, however, as participants studied faces that did not match any of the faces in the lineup. The results of the study showed that the lineup suspects were chosen at higher than chance levels, suggesting that picking foils that are similar to the suspect but not to each other might make the suspect appear distinctive. Subsequent investigations using real faces found that if the foils were chosen to match the suspect, matching the foils to each other reduced the rate of correctly guessing the suspect (Wogalter, Marwitz, & Leonard, 1992). These findings suggest that participants are able to determine which member is the focal point of the lineup based on examining the distribution of feature matches across lineup members.

Finally, the applied implications of adopting sequential over simultaneous lineups hinges on how often guilty as opposed to innocent suspects appear in police lineups. Consistent with previous work, we found that by switching from simultaneous to sequential lineups, hits are reduced in addition to lowering the rate at which innocent suspects are identified. If most often lineups contain guilty culprits, then sequential lineups will decrease the odds that they are positively identified by eyewitnesses. This point may seem moot in a justice system modeled after Blackstone's 10:1 ratio (Volokh, 1997), wherein it is "Better that ten guilty persons escape than that one innocent suffer." On the other hand, "Better for whom?"

APPENDICES

Table 1

Identification Responses by Lineup Construction, Identification Procedure, and Target Level for Experiment 1

Lineup Construction and ID Procedure	Target	Suspect	Foil	Reject
Matched:				
Sequential	Identical	0.37 (.03)	0.27 (.03)	0.36 (.03)
	Similar	0.13 (.03)	0.30 (.03)	0.57 (.03)
Simultaneous	Identical	0.45 (.03)	0.26 (.03)	0.29 (.03)
	Similar	0.24 (.03)	0.32 (.03)	0.44 (.03)
Random:				
Sequential	Identical	0.33 (.05)	0.30 (.04)	0.37 (.04)
	Similar	0.24 (.05)	0.36 (.04)	0.40 (.05)
Simultaneous	Identical	0.52 (.05)	0.23 (.04)	0.25 (.04)
	Similar	0.34 (.05)	0.27 (.04)	0.39 (.05)

Note: Data shown are the mean rate (SD) of suspect picks, foil picks, and lineup rejections.

Table 2

Mean Rate of Target Identifications for Each of the Study Faces in the Matched Condition in Experiment 1 by the Facial Feature Used to Match the Foils

	Identical Target				**Feature Removed Target**			
Study Face	Eye	Mouth	Face	Overall Mean	Eye	Mouth	Face	Overall Mean
1	0.53	0.64	0.88	0.68	0.47	0.26	0.13	0.29
2	0.32	0.53	0.44	0.43	0.36	0.33	0.08	0.26
3	0.55	0.63	0.58	0.59	0.50	0.33	0.33	0.39
4	0.30	0.35	0.58	0.41	0.22	0.60	0.47	0.43
5	0.75	0.79	0.89	0.81	0.05	0.25	0.08	0.13
6	0.68	0.94	0.94	0.85	0.33	0.43	0.90	0.55
7	0.38	0.33	0.31	0.34	0.44	0.28	0.18	0.30
8	0.27	0.73	0.60	0.53	0.00	0.17	0.94	0.37
9	0.65	0.57	0.62	0.61	0.11	0.75	0.38	0.41
10	0.76	0.67	0.68	0.70	0.50	0.63	0.71	0.61
11	0.84	0.75	0.83	0.81	0.27	0.29	0.67	0.41
12	0.79	0.50	0.50	0.60	0.17	0.71	0.33	0.40
Mean	0.57	0.62	0.65	0.61	0.29	0.42	0.43	0.38
SD	0.21	0.18	0.20	0.17	0.18	0.20	0.31	0.13

Table 3

Identification Responses for Full Lineups (Experiment 1) and Target Removed (Experiment 2) Lineups in the Matched Condition by Identification Procedure

		Full Lineup				Removed Lineup			Difference (Removed-Present)		
Study Face	Procedure	Target	Popular	Other	Reject	Popular	Other	Reject	Popular	Other	Reject
1*	Seq	0.25	0.26	0.11	0.38	0.19	0.34	0.47	-0.06	0.08	0.36
	Sim	0.42	0.16	0.22	0.20	0.22	0.47	0.31	-0.20	0.31	0.09
2	Seq	0.12	0.10	0.24	0.54	0.16	0.33	0.51	0.04	0.23	0.27
	Sim	0.31	0.21	0.24	0.24	0.21	0.53	0.26	-0.10	0.32	0.02
3*	Seq	0.28	0.15	0.11	0.46	0.23	0.25	0.52	-0.05	0.10	0.41
	Sim	0.30	0.13	0.16	0.41	0.32	0.44	0.25	0.02	0.31	0.09
4*	Seq	0.16	0.22	0.22	0.40	0.38	0.24	0.38	0.22	0.02	0.16
	Sim	0.38	0.16	0.22	0.24	0.23	0.54	0.23	-0.15	0.38	0.01
5*	Seq	0.28	0.19	0.12	0.41	0.35	0.15	0.50	0.07	-0.04	0.38
	Sim	0.31	0.14	0.19	0.36	0.22	0.39	0.39	-0.09	0.25	0.20
6	Seq	0.32	0.09	0.09	0.50	0.13	0.24	0.63	-0.19	0.15	0.54
	Sim	0.54	0.05	0.08	0.33	0.14	0.36	0.50	-0.40	0.31	0.42
7	Seq	0.11	0.23	0.19	0.47	0.29	0.29	0.42	0.18	0.06	0.23
	Sim	0.26	0.16	0.24	0.34	0.27	0.47	0.26	0.01	0.31	0.02
8	Seq	0.11	0.15	0.14	0.60	0.17	0.35	0.48	0.06	0.20	0.34
	Sim	0.41	0.13	0.10	0.36	0.15	0.38	0.47	-0.26	0.25	0.37
9*	Seq	0.34	0.19	0.10	0.37	0.28	0.26	0.46	-0.06	0.07	0.36
	Sim	0.28	0.19	0.06	0.47	0.50	0.28	0.22	0.22	0.09	0.16
10	Seq	0.34	0.05	0.05	0.56	0.18	0.33	0.49	-0.16	0.28	0.44
	Sim	0.28	0.06	0.15	0.51	0.07	0.59	0.34	-0.21	0.53	0.19
11*	Seq	0.44	0.10	0.06	0.40	0.10	0.32	0.58	-0.34	0.22	0.52
	Sim	0.30	0.10	0.17	0.43	0.15	0.39	0.46	-0.15	0.29	0.29
12	Seq	0.19	0.16	0.17	0.48	0.17	0.22	0.61	-0.02	0.06	0.44
	Sim	0.32	0.12	0.10	0.46	0.09	0.60	0.31	-0.23	0.48	0.21

*Indicates most popular foil in the simultaneous condition was also the most popular foil within the sequential condition.

Table 4

Identification Responses for Full Lineups (Experiment 1) and Target Removed Lineups (Experiment 2) in the Random Condition by Identification Procedure

		Full Lineup				Removed Lineup			Difference (Removed-Present)		
Study Face	Procedure	Target	Popular	Other	Reject	Popular	Other	Reject	Popular	Other	Reject
1*	Seq	0.21	0.21	0.10	0.48	0.08	0.15	0.77	-0.13	0.05	0.29
	Sim	0.62	0.10	0.11	0.17	0.13	0.59	0.28	0.03	0.48	0.11
2	Seq	0.25	none	0.31	0.44	none	0.66	0.34	none	0.35	-0.10
	Sim	0.26	none	0.33	0.41	none	0.77	0.23	none	0.44	-0.18
3	Seq	0.28	0.12	0.12	0.48	0.19	0.43	0.38	0.07	0.31	-0.10
	Sim	0.38	none	0.20	0.42	none	0.57	0.43	none	0.37	0.01
4*	Seq	0.19	0.44	0.17	0.20	0.32	0.32	0.36	-0.12	0.15	0.16
	Sim	0.69	0.11	0.05	0.15	0.28	0.49	0.23	0.17	0.44	0.08
5	Seq	0.28	0.11	0.00	0.61	0.25	0.24	0.51	0.14	0.24	-0.10
	Sim	0.35	none	0.19	0.46	none	0.36	0.64	none	0.17	0.18
6	Seq	0.40	0.08	0.04	0.48	0.17	0.47	0.36	0.09	0.43	-0.12
	Sim	0.46	none	0.23	0.31	none	0.77	0.23	none	0.54	-0.08
7*	Seq	0.24	0.28	0.16	0.32	0.36	0.19	0.45	0.08	0.03	0.13
	Sim	0.42	0.27	0.16	0.15	0.38	0.31	0.31	0.11	0.15	0.16
8	Seq	0.32	0.11	0.25	0.32	0.13	0.59	0.28	0.02	0.34	-0.04
	Sim	0.50	0.08	0.11	0.31	0.23	0.51	0.26	0.15	0.40	-0.05
9	Seq	0.25	0.17	0.12	0.46	0.13	0.36	0.51	-0.04	0.24	0.05
	Sim	0.31	0.11	0.20	0.38	0.10	0.28	0.62	-0.01	0.08	0.24
10*	Seq	0.12	0.23	0.37	0.28	0.19	0.58	0.23	-0.04	0.21	-0.05
	Sim	0.31	0.19	0.19	0.31	0.18	0.69	0.13	-0.01	0.50	-0.18
11	Seq	0.24	0.28	0.24	0.24	0.04	0.36	0.60	-0.24	0.12	0.36
	Sim	0.26	none	0.26	0.48	none	0.54	0.46	none	0.28	-0.02
12	Seq	0.52	none	0.20	0.28	none	0.52	0.48	none	0.32	0.20
	Sim	0.50	none	0.35	0.15	none	0.77	0.23	none	0.42	0.08

*Indicates most popular foil in the simultaneous condition was also the most popular foil within the sequential condition.

Table 5

Correlations of Identification Dependent Measures with Lineup Fairness Measures Derived from Simultaneous and Sequential Mock Witness Task in Experiment 3.

ID Procedure and Target Level	Outcome Measure	Acceptable Lineup Members	Effective Size	Functional Size
Sequential Identical Target	Choosing	-0.18	-0.17	-0.18
	Accuracy	-0.19	-0.18	0.29
	Target/ID	-0.01	-0.01	0.24
Sequential Feature Removed	Choosing	-0.39*	-0.39*	0.31
	Accuracy	-0.22	-0.22	0.15
	Target/ID	0.27	-0.21	0.27
Simultaneous Identical Target	Choosing	0.02	0.15	0.03
	Accuracy	0.27	0.22	0.07
	Target/ID	0.27	0.29	0.03
Simultaneous Feature Removed	Choosing	-0.49*	-0.26	-0.04
	Accuracy	-0.22	-0.26	-0.06
	Target/ID	0.07	-0.06	-0.08

* Indicates $p<.05$

Note: Results are reported separately for each identification procedure by target level condition.

Table 6

Correlations of Similarity Measures Obtained in Experiment 3 with Target Identifications from the Full Simultaneous and Sequential Lineups in Experiment 1

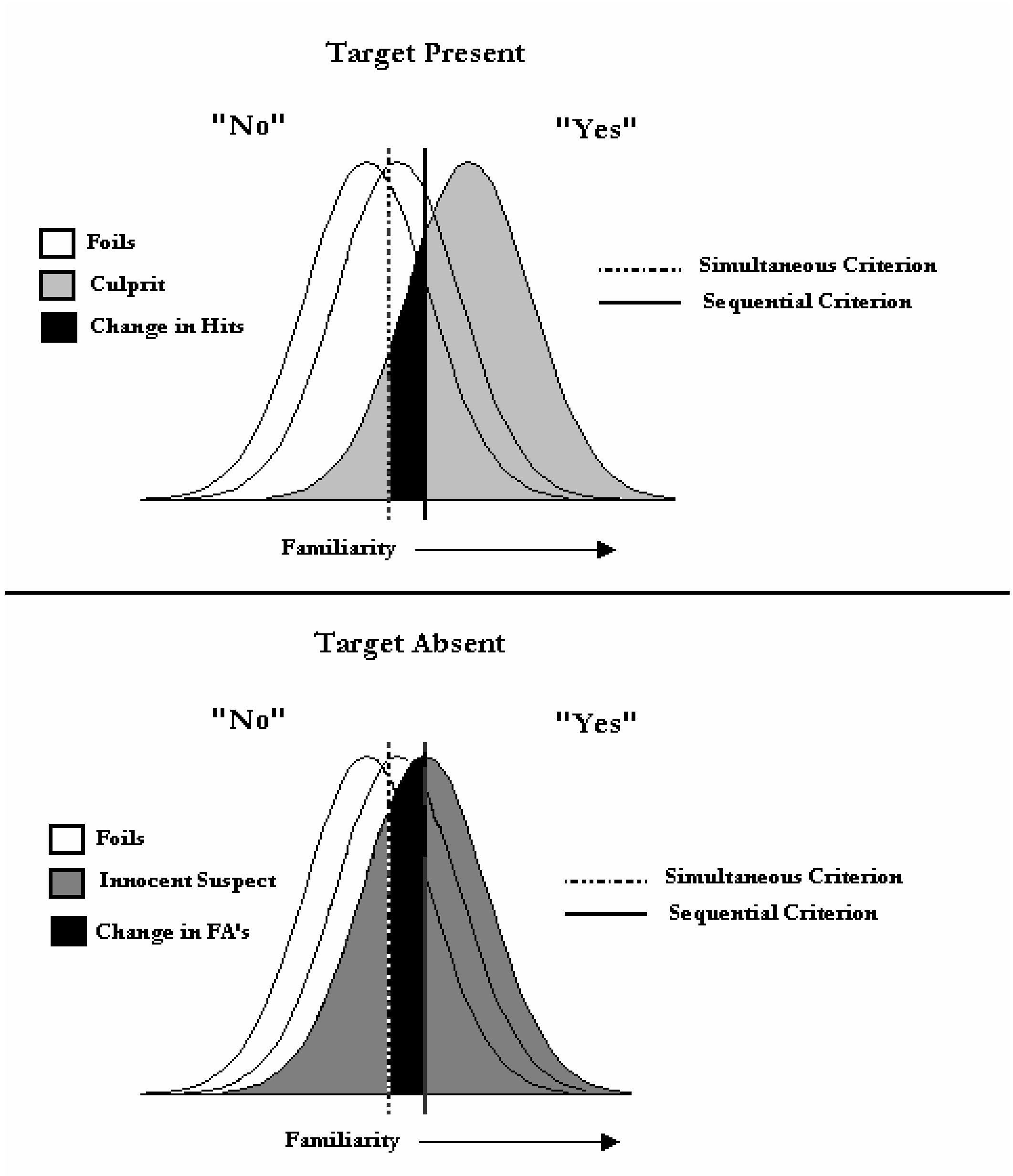
		Sequential		**Simultaneous**	
		Identical	Removed	Identical	Removed
Overall Target Choice Rate	Global Sim	-0.21*	-0.27*	-0.22	-0.08
	Global Seq	-0.22	-0.03	-0.26	0.29*
	Suspect to Foils	0.09	-0.15	0.08	0.18
	Foil to Foil	0.00	-0.36	0.77	-0.44*
	S-F	0.18	0.04	.0.24*	0.39*
	All Members	0.00	-0.35*	0.02	-0.41*
	Variability	-0.41	-0.26	0.08	-0.48*
Target/Choice Rate	Global Sim	-0.18	-0.11	-0.13	-0.05
	Global Seq	-0.24	-0.13	0.00	0.22
	Suspect to Foils	0.05	0.03	0.14	-0.19
	Foil to Foil	0.01	-0.22	0.10	-0.51*
	S-F	0.07	0.12	0.09	.0.49*
	All Members	0.02	-0.20	0.11	.-0.48*
	Variability	0.00	-0.28	0.03	.-0.52*

*Indicates $p<.05$.

Table 7

Identification Responses in Experiment 1 by Identification Procedure, and Target Level Compared to Showup (Seblay et al., 2003) and Lineup (Steblay et al., 2001) Meta-Analyses

	Suspect	Target	Foil	Reject
Experiment 1				
Simultaneous	Identical	0.49	0.25	0.27
	Similar	0.29	0.30	0.42
Sequential	Identical	0.35	0.34	0.37
	Similar	0.19	0.33	0.49
Showups	Identical	0.59		0.41
	Similar	0.28		0.72
Meta-Analyses				
Simultaneous	Identical	0.50	0.24	0.26
	Similar	0.27	0.24	0.49
Sequential	Identical	0.35	0.19	0.46
	Similar	0.09	0.19	0.72
Showups	Identical	0.47		0.53
	Similar	0.15		0.85



<u>Figure 1</u>

Signal detection representation of the change in hits and false alarms expected in moving from a simultaneous to a sequential procedure owing to higher criterion placement in the sequential compared to simultaneous procedure.

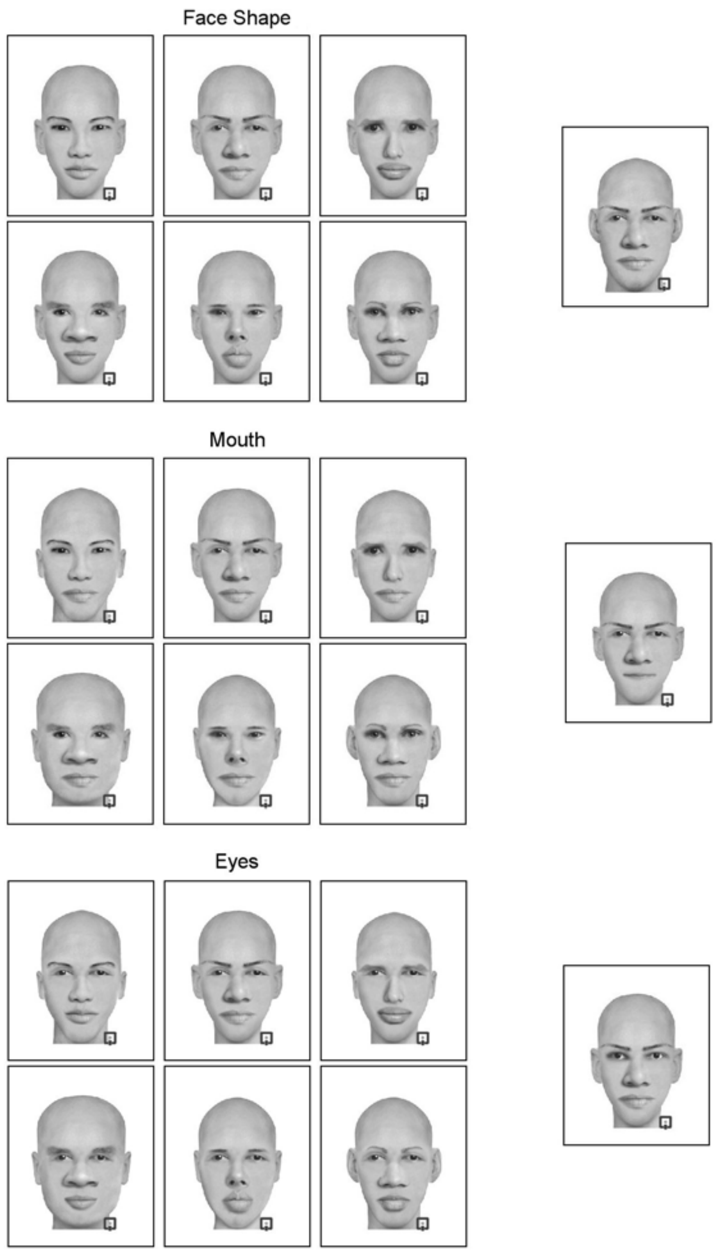
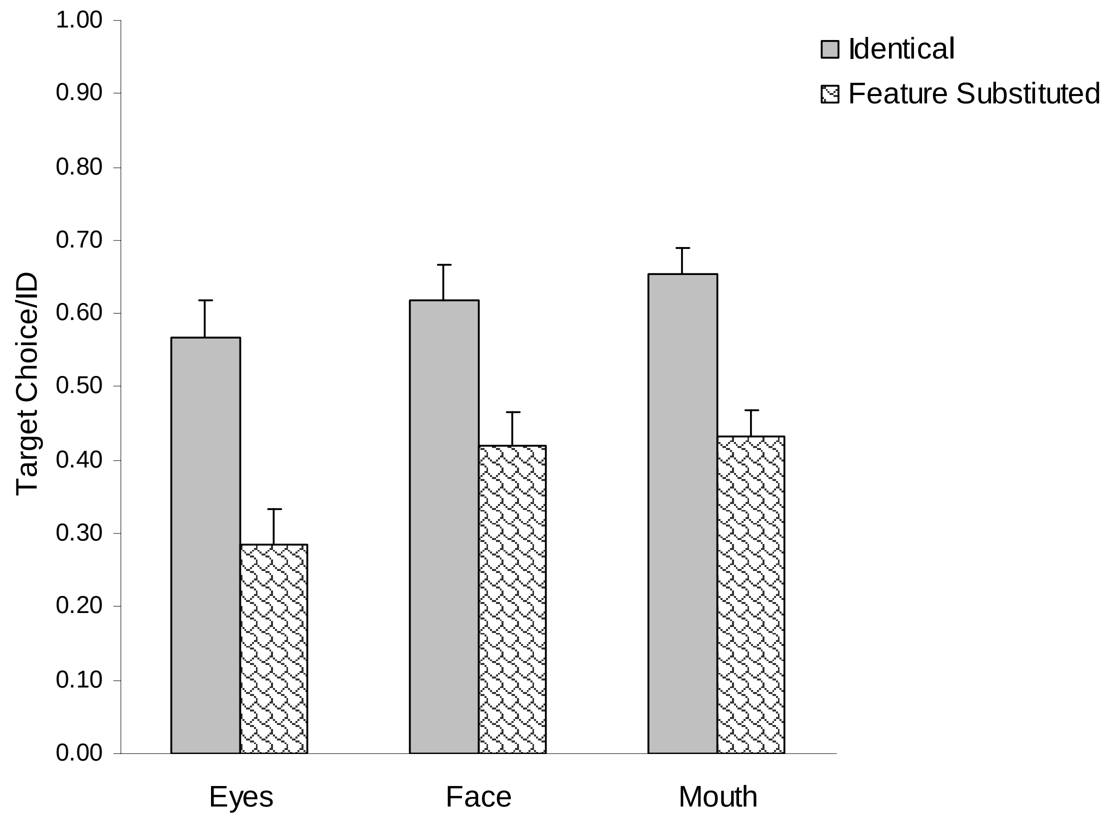


Figure 2

Example of lineup stimuli for a study face (row 1, center) in the matched condition for each feature level, which is indicated at the top of the lineup. The feature substituted target that took the place of the study face is located to the right of each lineup.



<u>Figure 3</u>

Average rate of target identifications (+1 SE) for choosers by target level and feature condition in Experiment 1.

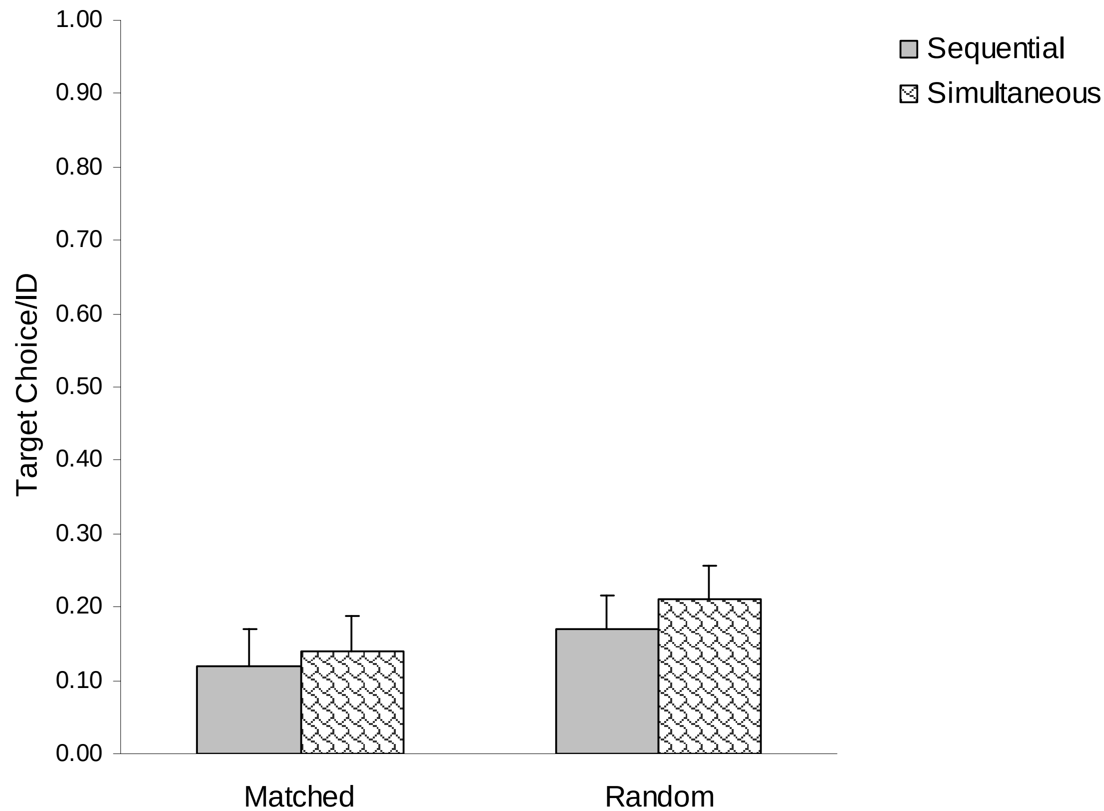
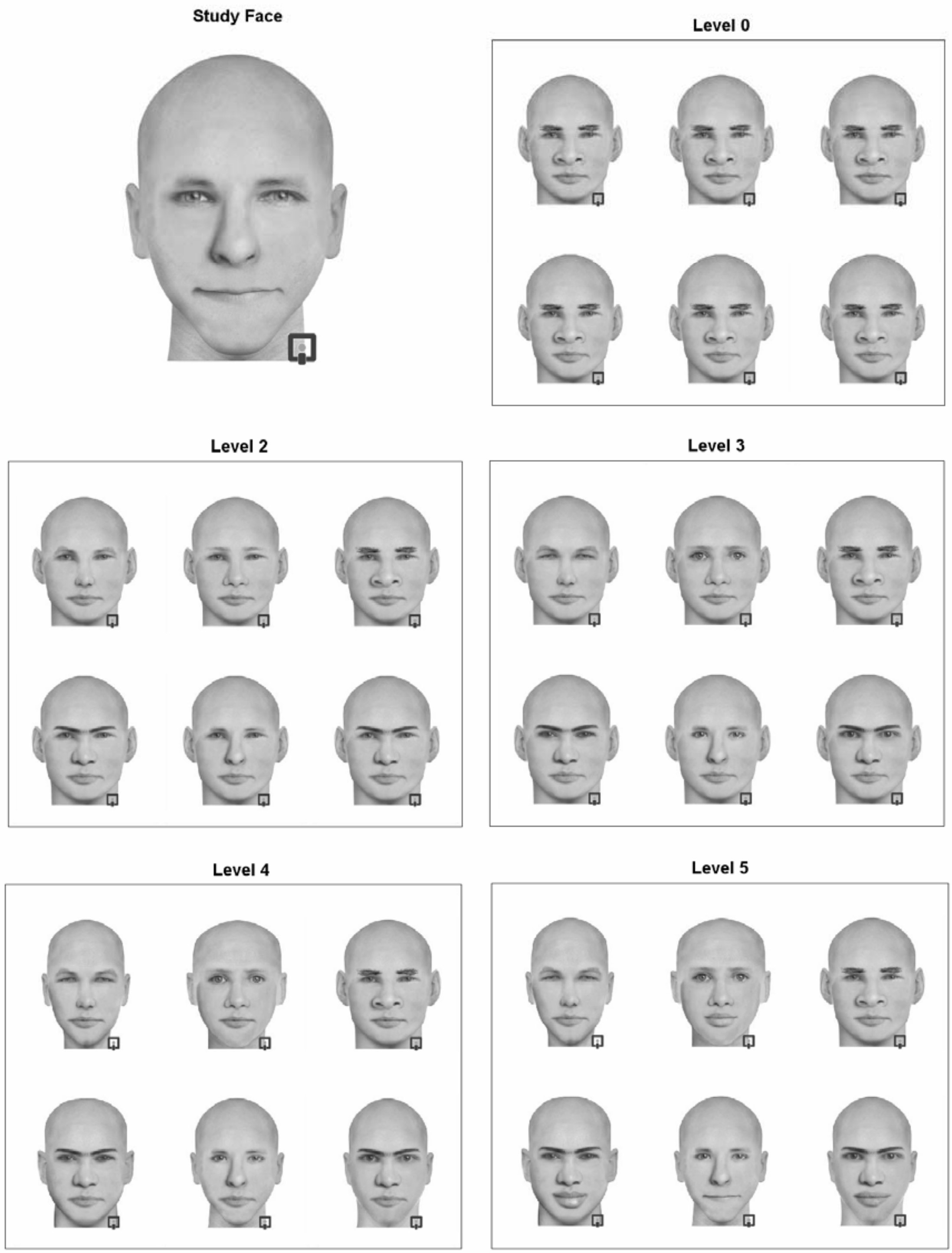


Figure 4

Average rate of mock witness target identifications (+1 SE) by lineup similarity condition in Experiment 3.

Figure 5

Illustration of the multiple lineup procedure used in Experiment 4. A complete lineup set for one study face is displayed. As participants progressed through the lineup levels (0-5, which correspond to the number of features that differ among the members in the lineup), the features of the faces in the lineup increasingly varied across faces. The goal of the participant was to accurately identify the study face at the lowest feature level possible.


Study Face
Level 0
Level 2
Level 3
Level 4
Level 5

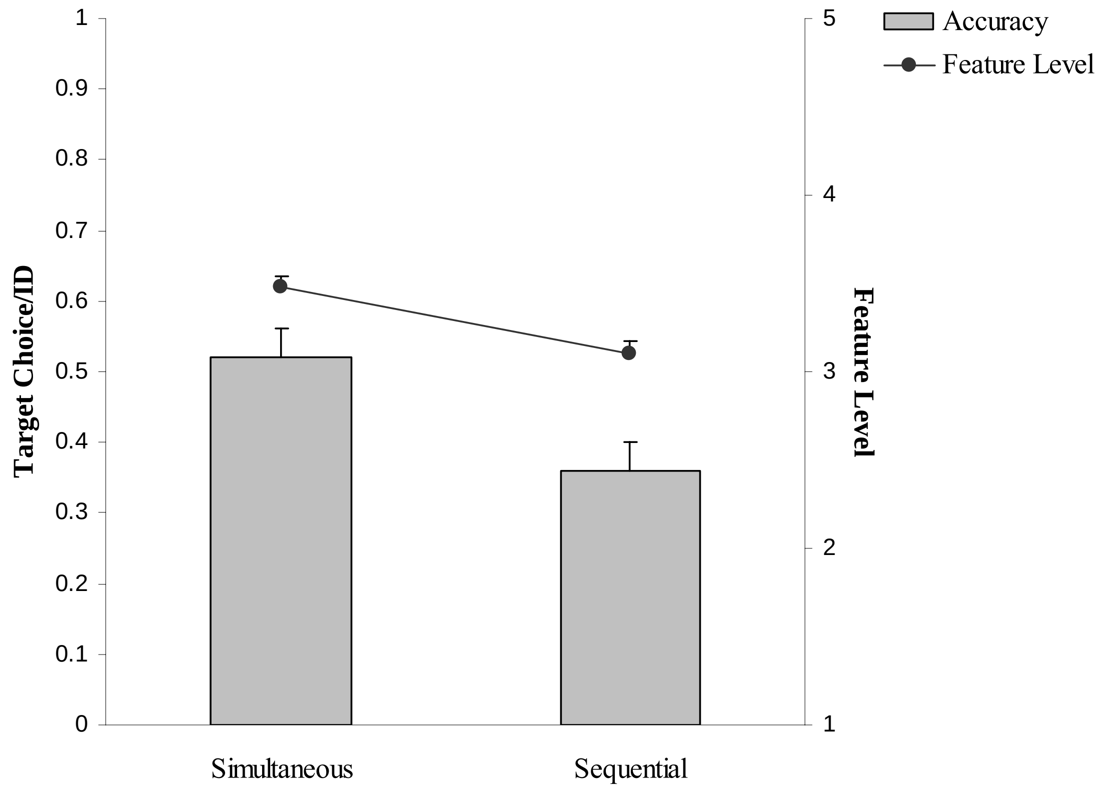


Figure 6

Average accuracy rate and average feature level requested (+1/2 SE) by identification procedure in Experiment 4.

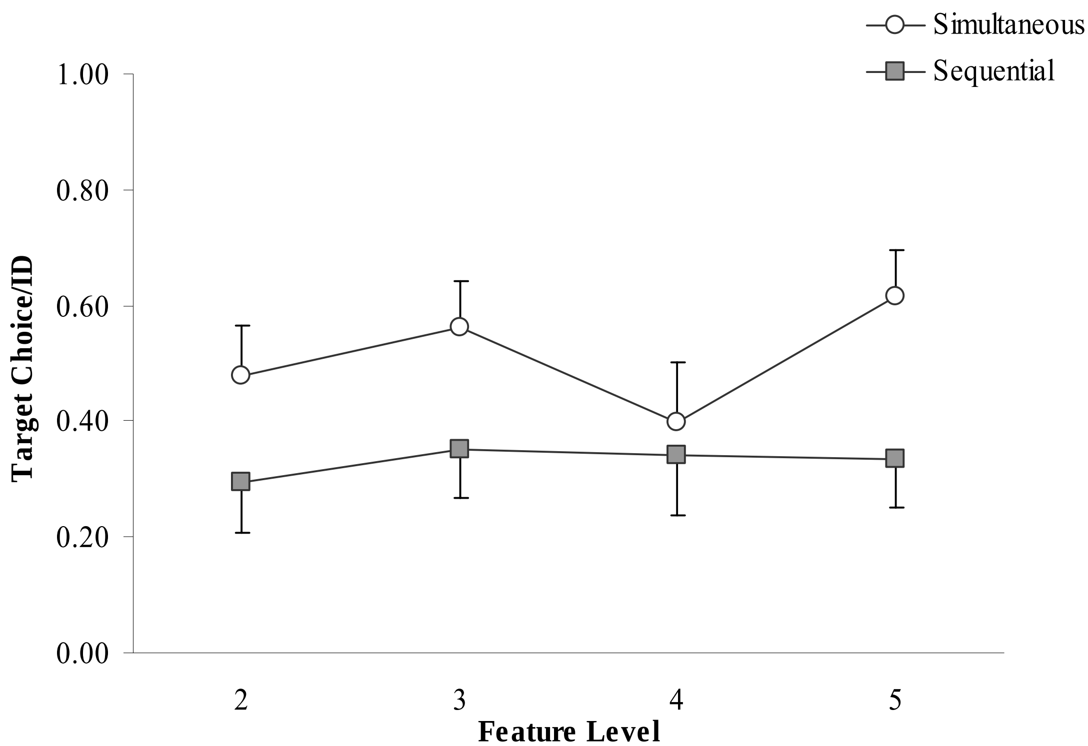


Figure 7

Average rate of identifying target (+1 SE) by level at which participants made a choice and identification procedure in Experiment 4.

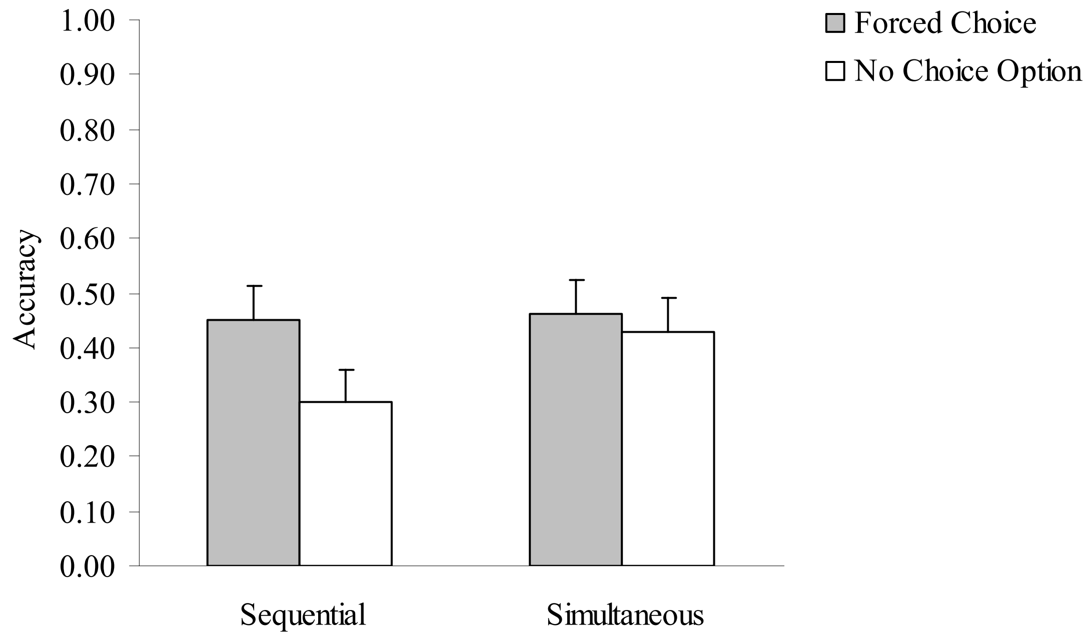


Figure 8

Average rate of choosing the suspect (+1 SE) collapsed across feature level by choice level and identification condition in Experiment 5.

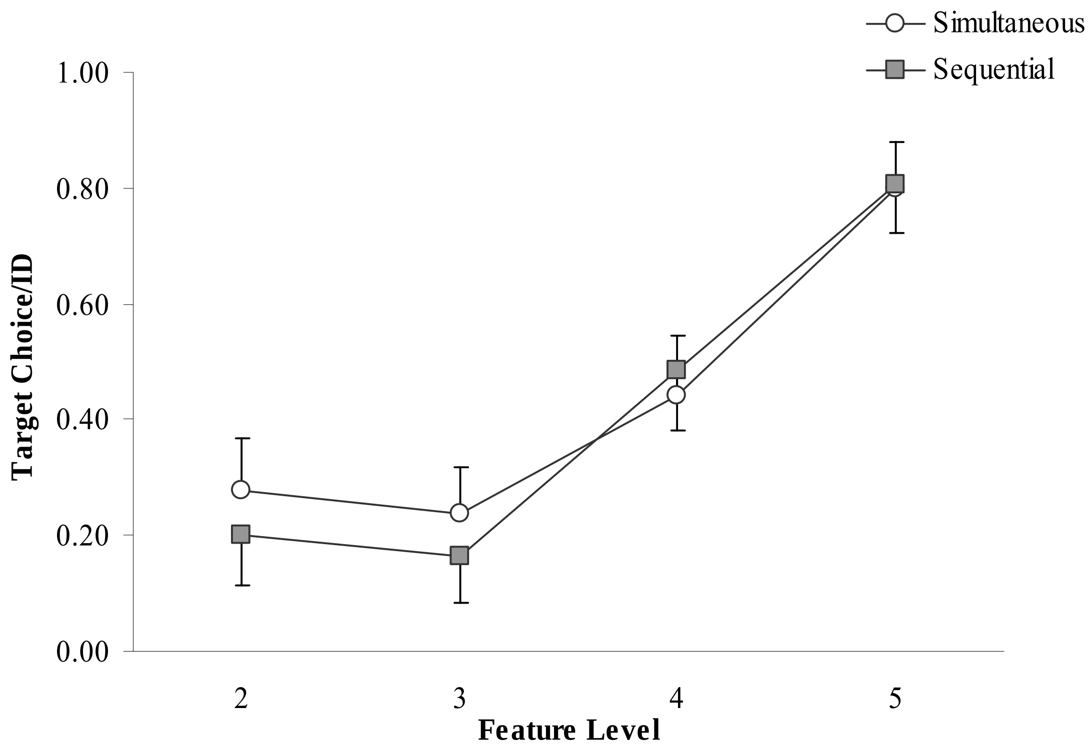


Figure 9

Average rate of identifying the target for choosers by feature level and lineup procedure in Experiment 5. The number of features participants could use in identifying the target was fixed.

REFERENCES

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. Law & Human Behavior, 25(5), 475-491.

Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence-accuracy relationship in eyewitness identifications. Basic and Applied Social Psychology, 11(1), 101-115.

Brigham, J. C., Maass, A., Snyder, L. D., Spaulding, K. B. (1982). Accuracy of eyewitness identification in a field setting. Journal of Personality and Social Psychology, 42(4), 673-681

Clark, S. E. (2003). A memory and decision model for eyewitness identification. Applied Cognitive Psychology, 17(6), 629-654. Retrieved August 8, 2005, from PsycINFO (1840-Current) database.

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. Law &Human Behavior, 29(2), 151-172

Clark, S., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. Law & Human Behavior, 25(3), 199-216.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? The view from UCSD. In E. Michael J. Wenger & E. James T. Townsend (Eds.), Computational, geometric, and process perspectives on facial cognition: Contexts and challenges. (pp. ix, 514): Lawrence Erlbaum Associates, Inc., Publishers.

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.

Doob, A. N., & Kirshenbaum, H. (1973). Bias in police lineups - Partial remembering. Journal of Police Science and Administration, 1, 287-293.

Ebbesen, E. B., & Flowe, H. D. (2001). Simultaneous v. Sequential Lineups: What Do We Really Know? Unpublished manuscript.

Ebbesen, E. B., & Wixted, J. (1996). A signal detection analysis of the relationship between confidence and accuracy in eyewitness memory. Unpublished manuscript.

Farmer, J. (2001). http://www.state.nj.us/lps/dcj/agguide/photoid.pdf Retrieved August 8, 2005

Flowe, H. D., Ebbesen, E. B., Burke, C., & Chivabundtt, P. (2001). At the scene of the crime: An examination of the external validity of published studies on line-up identification accuracy. Paper presented at the American Psychological Society, Toronto, Canada.

Gronlund, S. D. (2004). Sequential lineups: Shift in criterion or decision strategy? Journal of Applied Psychology, 89(2), 362-368.

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. Applied Cognitive Psychology, 19(1), 23-37.

Haig, N. D. (1984). The effect of feature displacement on face recognition. Perception, 13(5), 505-512.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and Sequential Lineups: Decision processes of accurate and inaccurate witnesses. Applied Cognitive Psychology, 15, 659-71

Laughery, K. R., Jensen, D. G., and Wogalter, M. S. (1988). Response bias with prototypic faces. In M. M. Gruneberg, R. Sykes and P. Morris (eds.), Practical aspects of memory: Current research and issues (pp. 157-162). Chichester: Wiley.

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

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. Journal of Applied Psychology, 84(3), 315-321.

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

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. Law & Human Behavior, 21, 404-411.

Lindsay, R. C. L., Smith, S. M., & Pryke, S. (1999). Measures of lineup fairness: Do they postdict identification accuracy? Applied Cognitive Psychology, 13(SI), S93-S107.

Lindsay, R. C., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy. Canadian Journal of Behavioural Science, 19(4), 463-478.

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.

Luus, C. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. Law & Human Behavior, 15(1), 43-57.

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. Law &Human Behavior, 5(4), 299-309.

Malpass, R. S., & Devine, P. G. (1983). Measuring the fairness of eyewitness identification lineups. In S. M. A. Lloyd-Bostock and B. R. Clifford (eds.), Evaluating witness evidence (pp. 81-102). London: Wiley.

Malpass, R. S. & Lindsay, R. C. L. (1999). Measuring lineup fairness. Applied Cognitive Psychology, 13(SI), S1-S7.

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

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

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. British Journal of Psychology, 80(3), 313-331.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? Journal of Applied Psychology, 85,542-550.

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

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

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. Law & Human Behavior, 25(5), 459-473.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. Law & Human Behavior, 27(5), 523-540.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 24(6), 1379-1396.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 46(2), 225-245.

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

Volokh, A. (1997). n guilty men. University of Pennsylvania Law Review, 146, 173.

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., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. Law &Human Behavior, 3(4), 285-293.

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), 603-647.

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. Psychological review, 107(2), 368-376.

Wogalter, M. S., Marwitz, D. B., & Leonard, D. C. (1992). Suggestiveness in photospread line-ups: Similarity induces distinctiveness. Applied Cognitive Psychology, 6(5), 443-453.