Title

Decision Processes in Eyewitness Identification

Author

Moreland, Molly Bettis

Publication Date

2015

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Decision Processes in Eyewitness Identification

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Molly Bettis Moreland

August 2015

Dissertation Committee:

Dr Steven Clark, Chairperson
Dr Curt Burgess
Dr Aaron Seitz
Dr Robert Rosenthal

The Dissertation of Molly Bettis Moreland is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I thank my Dissertation Chair and advisor, Steve Clark, who "adopted" me at a critical point in my graduate education. Steve's support, extensive knowledge, and incredible patience have been fundamental to my work. I have learned so much about conducting research from Steve–from looking for the devil in the details, to deciphering the "yummy data." He is an outstanding advisor.

I thank the members of my dissertation committee, Aaron Seitz, Bob Rosenthal, and Curt Burgess, for invaluable feedback. Aaron taught me to program in MATLAB–an essential skill that made Experiment 3 (and future experiments) possible. Bob's careful proofreading of this dissertation was extremely helpful, and his suggestions for statistical analyses will undoubtedly prove useful. Curt has served on all of my committees throughout my time in graduate school–his feedback, questions, and thoughtful discussions have always benefited my work.

I am grateful for the financial support of the Presley Center for Crime and Justice Studies for two graduate fellowships and the UC Riverside Graduate Division for the Dissertation Research Grant.

I thank Mike Erickson, who accepted me into the psychology Ph.D. program at UCR and provided me with an excellent first two years of graduate school. I owe my data analytic and programming skills to Mike. His instruction in statistics, R, and LaTeX have shaped my approach to working with and presenting data, and his mentoring shaped me as a researcher.

I gratefully acknowledge my labmates, Rakel Larson, Ryan Rush, Marie Hicks, Pat LaShell, Elizabeth Melton, and especially Justin Estep–who encouraged and inspired me to go to graduate school. I also am indebted to many undergraduate research assistants who collected lots and lots of data for me over the years.

I thank my parents, Karen Moreland–for being an energetic and motivating force of nature and teaching me how to write well–and Marc Moreland–for always being gentle, kind,

and encouraging, and teaching me that “it’s not time to worry, yet.” Their unconditional love and patience has always comforted and amazed me.

I thank my husband, John Dusel, whose love and support is matchless, and my brother, Chris Moreland, whose humor and cynicism is indispensable. I’ve loved sharing the graduate experience with you both. Finally, I thank August Moreland Dusel, for showing me that change is beautiful and ever-present.

To John and August–the journey is the destination.

ABSTRACT OF THE DISSERTATION

Decision Processes in Eyewitness Identification

by

Molly Bettis Moreland

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, August 2015
Dr Steven Clark, Chairperson

The dominant theory of decision-making in eyewitness identification, based on a distinction between absolute and relative judgments, assumes that relative judgments (identifying the best match relative to the other lineup members) increases identification errors (Wells, 1984). This distinction also assumes that comparisons among lineup members underlying relative judgments increases errors, as evidenced by the sequential lineup advantage (Lindsay & Wells, 1985). Sequential lineups preclude comparisons among lineup members by presenting lineup members separately. Recently, and in contrast to the dominant view, receiver operating characteristic (ROC) studies found a simultaneous accuracy advantage (Mickes, Flowe, & Wixted, 2012) that may be due to underlying comparative processes (J. T. Wixted & Mickes, 2014). This dissertation focuses on the effect of comparative strategies on eyewitness identification accuracy. Clark, Erickson, and Breneman (2011) simulated identification response probabilities for different decision rules and foil selection methods using the computational model, WITNESS (Clark, 2003). These simulations showed a disadvantage for relative judgment strategies for lineups with foils matched to the appearance of the suspect. Experiment 1 provides an empirical test of the WITNESS model's predictions. Participants viewed lineups that differed in suspect guilt or innocence and foil selection method. The innocent suspect either appeared in different-foils lineups, the foil selection method used by police, or same-foils lineups, a method used in eyewitness identification research that holds foils constant across lineup type. Additionally, lineup instructions either directed witnesses to use an absolute or relative judgment decision rule. ROC curves indicated a small advantage for relative decision rules for different-foils and same-foils lineups. Contrary to the dominant view, the results suggest that relative judgments (and comparisons among lineup members) may improve accuracy. Experiment 2 investigated comparative processes independently of decision rule. Participants viewed lineups that differed in arrangement. Similar foils surrounded the suspect in the high-similarity neighbor condition, designed to facilitate comparative processes, and dissimilar foils surrounded the suspect in the low-similarity neighbor condition. Results showed an accuracy advantage for high-similarity neighbors based on ROC analyses. Experiment 3 investigated the degree to which comparisons among lineup members underlie the simultaneous accuracy advantage. Witnesses viewed lineups differing in lineup presentation (simultaneous or sequential), and neighboring foil similarity. Results showed a larger accuracy advantage of simultaneous high-similarity neighbor lineups. High-similarity neighbors may facilitate comparative strategies in simultaneous lineups but not sequential lineups, corroborating recent reports of a simultaneous advantage and contrary to the widespread "sequential superiority" view.

Contents

List of Tables

List of Figures

Chapter 1: Eyewitness Decision Making

> You ever read descriptions by eyewitnesses? They're nuts. My cop friend says sometimes when they'd plant him in the lineup, he got picked out over and over again. People swore their eyes out he did whatever it was.
>
> John Steinbeck, *The Winter of Our Discontent*

The Problem of Eyewitness Identification

Eyewitnesses make identification errors. According to one estimate, 75% of the 329 false convictions cleared by DNA evidence in the United States involved a particular eyewitness error, a false identification of an innocent person (*http://www.innocenceproject.org/*). The percentage of known false convictions involving false identification of the innocent varies depending on inclusion criteria. For example, Gross and Shaffer (2012) included only witnesses who made a false identification based on a memory error and excluded witnesses who lied by making deliberate misidentifications or fabricating crimes. They found that 43% of 873 DNA and non-DNA exoneration cases from 1989 to 2012 involved false identifications. Including witnesses who lied with those who made a false identification increases the estimate of exoneration cases involving eyewitness identification errors to 76%. According to Gross and Shaffer, false identifications based on eyewitness memory errors tend to be most prevalent in robbery cases (81%), sexual assault cases (80%), and other non-homicide violent crimes (51%) and less prevalent in homicides (27%), child sex abuse (26%), and nonviolent crimes (19%).

Some (but not all) factors commonly implicated in false identification exoneration cases are suggestive lineup procedures, multiple eyewitnesses, the innocent suspect's appearance, and cross-racial identifications. A case that exemplifies the latter two factors began in 1984. A black man, Ronald Cotton, was accused and subsequently convicted of raping a white woman, Jennifer Thompson. Cotton became a suspect in the Thompson case because of his resemblance to a composite drawing of the rapist displayed on television (Thompson-Cannino, Cotton, & Torneo, 2009). Cotton spent over a decade in prison and was exonerated based on DNA evidence in 1995. Bobby Poole, jailed for a similar crime in the same prison as Cotton, was Thompson's rapist. Cases like

Ronald Cotton's illustrate the potentially devastating consequences of false identification errors. False identifications, however, are not the only errors made by eyewitnesses. Other types of eyewitness errors have received less attention, possibly because these errors do not carry the risk of sending an innocent person to prison. For example, an identification of John Steinbeck's "cop friend"—a foil, known to be innocent—may absolve an innocent suspect or allow the perpetrator to go free. Rejections of a lineup (an indication that the person who committed the crime is not in the lineup) have similarly variable consequences depending on the guilt or innocence of the suspect. Consider a case from 1974. Two women were abducted in July during the day from Lake Sammamish State Park in Washington. The day of the abduction, five witnesses reported a suspicious man with a sling requesting help unloading a sailboat from his car. One witness reported running away from the man when she realized there was no sailboat in his car. In September, the remains of the two abductees were found two miles from the park. In November 1975, one of the witnesses was shown a photo lineup. The witness rejected the lineup, unable to make an identification of the suspect, Ted Bundy. Bundy was a serial rapist and murderer who killed approximately 30 victims by 1978. The Bundy case also exemplifies the utility of a correct identification. Carol DaRonch was abducted on a separate occasion by Bundy. She managed to escape and made a positive identification of Bundy from a seven-person photo lineup in 1975. The DaRonch identification helped precipitate an investigation of Bundy.

How do eyewitnesses make correct and incorrect identification decisions? The question is difficult to answer using real cases because whether the person convicted of a crime is truly guilty or innocent is often unknown. This limitation is called the ground-truth problem. Experimental studies of eyewitness identification avoid the ground truth problem because the perpetrator's identity in an experiment is known to a certainty. This is because the experimenter creates lineups such that half of the lineups include the perpetrator, and the other half include an innocent suspect. The experiments described in this dissertation use the experimental methods detailed below.

Eyewitness Identification Lineups and Experimental Simulations

In a typical eyewitness identification experiment, naïve participants become witnesses to a live or videotaped staged crime. These witnesses are later shown a lineup that contains a guilty or an innocent suspect (denoted guilty-suspect lineups and innocent-suspect lineups, respectively) and a number of foils (other lineup members) who are known to be innocent. In experiments, guilty suspect lineups simulate a real-world situation in which the police lineup contains the perpetrator, whereas innocent suspect lineups simulate a real-world situation in which the police lineup contains an innocent person.

The witness' response is typically classified into one of three categories: an identification of the suspect, an identification of a foil, or a lineup rejection—a statement that the perpetrator is not in the lineup. These responses are detailed in Table 1. In a guilty-suspect lineup, the sole correct response is an identification of the suspect, called a correct identification. Lineup rejections and foil identifications are errors. In an innocent-suspect lineup, the sole correct response is a lineup rejection. Foil identifications and suspect identifications are both errors in innocent-suspect lineups, although an identification of an innocent suspect (false identification) is the only response that could result in wrongful incarceration.

Table 1
Possible responses to guilty-suspect and innocent-suspect lineups and response classifications. Computation of identification rates depending on response and lineup condition, assuming 100 guilty-suspect and 100 innocent-suspect lineups were presented.

Response	Guilty-Suspect Lineup	Innocent-Suspect Lineup
Suspect	Correct Identification $60/100 = .60$	False Identification $30/100 = .30$
Foil	Foil Identification $15/100 = .15$	Foil Identification $20/100 = .20$
Rejection	Incorrect Rejection $25/100 = .25$	Correct Rejection $50/100 = .50$

How are identification rates calculated from experimental data? Consider an experimental situation represented in Table 1 in which 100 witnesses view a guilty-suspect lineup, and 100 witnesses view an innocent suspect lineup. If 60 witnesses in the guilty-suspect condition identify the guilty suspect, the correct identification rate is $60/100 = .60$. If 30 of the witnesses in the innocent-suspect condition identify the innocent suspect, the false identification rate is $30/100 = .30$. Foil identifications and lineup rejections are calculated similarly in each lineup condition.

Theories of Eyewitness Identification Decisions

The absolute-relative distinction. How do eyewitnesses make memory decisions when faced with a criminal lineup? According to Wells' (1984) absolute-relative distinction, witnesses use one of two decision strategies when making identification decisions. Wells suggested that witnesses using a relative judgment process identify the lineup member who is the best match to memory relative to the other lineup members. Witnesses using an absolute judgment process, in contrast, identify a lineup member if the value of that member's match to memory exceeds a criterion.

The distinction between absolute and relative judgment processes was advanced as an explanation for a particular pattern of eyewitness identification results—witnesses were assumed to use absolute judgment processes when a lineup manipulation decreased the false identification rate without substantial decrease to the correct identification rate. In contrast, witnesses were assumed to use relative judgment processes when a lineup manipulation increased the false identification rate without substantial increase to the correct identification rate (Lindsay & Wells, 1985; Wells, 1984; Mansour, Lindsay, Brewer, & Munhall, 2009; Lindsay, Pozzulo, Craig, Lee, & Corber, 1997).

For proponents of the absolute-relative distinction, the strongest evidence supporting the distinction came from experiments comparing simultaneous and sequential lineups (Lindsay & Wells, 1985; Kneller, Memon, & Stevenage, 2001; Mansour et al., 2009; Lindsay et al., 1997). In a simultaneous lineup, all lineup members are presented at the same time, and the witness has the opportunity to examine all lineup members—and make comparisons between them—before making any response. In a sequential lineup, the lineup members are presented one-at-a-time, and the witness is required to respond to each lineup member before the next lineup member is presented. The rationale behind the development of the sequential lineup was that it would reduce witnesses' tendency to make relative judgments by limiting their ability to compare the lineup members to each other; thus, the sequential lineup would improve accuracy.

In the first published study comparing simultaneous and sequential lineups, Lindsay and Wells (1985) showed that sequential lineups decreased the false identification rate compared to simultaneous lineups (.17 *vs.* .43) with only a small decrease in the correct identification rate (.50 *vs.* .58), exactly the pattern assumed to underlie the absolute-relative distinction. A meta-analysis of simultaneous-sequential comparisons corroborated the large decrease in false identifications and relatively small decrease in correct identifications for sequential lineups, a finding that was dubbed “sequential superiority" (Steblay, Dysart, Fulero, & Lindsay, 2001). Comparative processes, thought to underlie relative judgments and performance on simultaneous lineups, were assumed to increase error and decrease the accuracy of eyewitness identification evidence overall (Dunning & Stern, 1994; Kassin, Tubb, Hosch, & Memon, 2001; Wells, 1984; Woocher, 1977). Further, the advantage for sequential lineups lead to the recommendation that law enforcement adopt the sequential lineup as the standard identification procedure (Wells, 2006; U. S. Department of Justice, Office of Justice Programs, National Institute of Justice, 1999). A few states (e.g., New Jersey, Wisconsin, North Carolina) and several cities, and local jurisdictions have switched from simultaneous to sequential lineups.

As Wells, Memon, and Penrod (2006) have noted, the absolute-relative distinction has “permeated the literature on lineups" (p. 61) primarily based on the sequential advantage. Despite the widespread acceptance of the absolute-relative explanation of decision making, the original definitions of absolute and relative judgment processes were not well-specified. Are absolute and relative judgments simply decision rules specifying when to make an identification? Does the original conceptualization of relative judgments involve a process to determine the best matching lineup member and a decision rule? How is the best-matching lineup member determined “relative to the other lineup members?"

Wells (1984) described relative judgments as “reliance on choosing the person who most resembles the criminal" (p. 93) and the distinguishing characteristic between absolute and relative judgments as: “there is nothing in the relative-judgment processing strategy that would lead a witness to reject the lineup (i.e., make a nonidentification response)" (p. 94). Framed this way, a relative judgment is a rule: if there is a best match to memory, (and there always is) identify that best match. This rule does not specify how the witness determines the best match.

However, the process of computing the best match is central to many definitions of relative judgments. For example, Wells (1984) clarifies why witnesses with poor memories of a crime might still identify a lineup member: “upon first narrowing the options to one lineup member via a relative judgment, the witness may engage in a constructive memory process of incorporating detailed characteristics of that lineup member into his or her recollection" (p. 95). The quote seems to imply that relative judgments involve a process of elimination that yields a best matching lineup member, rather than simply a decision rule to identify the best match.

More typically, the process suggested to underlie relative judgments is described as involving comparisons between lineup members. For example: “the relative judgment strategy is defined as comparisons among lineup members" (Mansour et al., 2009, p.1013); “eyewitnesses tend to compare lineup members to each other to determine which one most closely resembles the perpetrator relative to the others, a process called relative judgment" (Steblay, Dysart, & Wells, 2011, p.495); and “witnesses will compare all of the choices to each other looking for the most familiar person in the display" (Ebbesen & Flowe, 2002, p. 4). Relative judgments have been designated as a “comparative strategy" (Clark & Davey, 2005) or similarly, “the relative comparison process" (Flowe & Ebbesen, 2007), and “relative comparisons" (Wixted & Mickes, 2014).

As mentioned, the absolute-relative distinction was advanced as a theory of eyewitness decision making and an explanation for a particular pattern of data. A distinction, however, is not a theory. Moreover, the absolute-relative distinction is underdeveloped, and underspecified. What is the theoretical basis for the prediction that relative judgments increase error? Does this prediction arise from any theory? A computational model is necessary to answer these questions. Comparisons of model predictions and empirical data inform and guide theoretical development (Clark, 2008; Hintzman, 1993). A formal model can reveal inconsistencies between the assumptions of the model and the observed outcome, and in this way, models can aid in the understanding of the underlying processes. The next section describes a formal computational eyewitness identification model that clarifies and instantiates absolute and relative decision making.

The WITNESS Model. WITNESS is a version of a class of recognition memory model called global matching models (Clark, 2003). In general, global matching models represent information as vectors of elements. They compute the match between a test item vector and a memory vector (often by a dot product) to generate response probabilities that can be compared to data (Clark & Gronlund, 1996). WITNESS is specifically adapted to generate response probabilities to compare to eyewitness identification data.

WITNESS represents the perpetrator of a crime as a vector of elements, $\mathbf{P}$. The witness' memory of the perpetrator is a noisy copy of $\mathbf{P}$, denoted $\mathbf{M}$. Each lineup member is represented by a vector of features $\mathbf{L}_i$. The probability that a given feature of $\mathbf{P}$ will be represented accurately in memory $\mathbf{M}$ is a, and the probability that the feature will be represented inaccurately is $1-a$. Memory accuracy a varies depending on factors that affect memory, like opportunity to observe and retention interval.

In addition to memory accuracy, identification performance in WITNESS also depends on similarity relationships between the innocent suspect and the perpetrator ($s[S,P]$), the foils and the perpetrator ($s[F,P]$), and the foils and the suspect ($s[F,S]$). WITNESS represents the similarity parameters as overlap between the vectors (overlap between features).

WITNESS instantiates the process of matching to memory as a dot product that compares each $\mathbf{L}_i$ to $\mathbf{M}$, yielding N (number of lineup members) match values. WITNESS makes an identification if the memory evidence exceeds a criterion according to a decision rule. The memory evidence is a weighted sum, $W_A(\text{BEST}) + W_R(\text{BEST} - \text{NEXT})$, where BEST is the match value of the lineup member that is the best match to memory, and NEXT is the match value of the lineup member that is the second best match to memory. W_A and W_R are weights that reflect use of absolute and relative judgments processes that sum to 1.0.

The absolute decision rule is instantiated in WITNESS by setting $W_A = 1.0$ and $W_R = 0.0$. This puts all of the weight on the value of the BEST match, comparing only the best match value to the criterion, C, the standard for making an identification. If no member's match value exceeds C, then WITNESS rejects the lineup. In contrast, a relative decision rule is instantiated by setting $W_R = 1.0$ and $W_A = 0.0$, in which case WITNESS identifies a lineup member if the difference between best matching member and the second best matching member exceeds a difference criterion, C_{diff}. This framework formalizes the absolute-relative distinction, providing clear operational definitions and allowing for testable predictions.

Goodsell, Gronlund, and Carlson (2010) modeled response probabilities of simultaneous-sequential comparisons using WITNESS. Sequential lineups were modeled by testing each lineup member L_i alone and identifying the first lineup member that exceeded C. If no lineup member exceeded C, the lineup was rejected. Goodsell et al. found that WITNESS produced a small sequential advantage when C was set high for sequential lineups (conservative responding) and C was set low for simultaneous lineups (liberal responding). WITNESS was unable to produce a large sequential advantage (like that shown by Lindsay and Wells (1985)) without adjustments that (1) shifted from an absolute to a relative decision rule after rejecting several foils and (2) improved WITNESS's memory comparison process over the course of the lineup. Rather than manipulating the decision weights W_A and W_R, the first adjustment was made by shortening the memory vector for relative judgments. Conceptually, shortening the memory vector simulates Wells' (1984) hypothesis that people using relative judgments probe memory less thoroughly. The difficulty simulating a large sequential advantage is suggestive that the model is wrong, or that the large sequential advantage may be spurious.

In concert with the latter possibility, the following sections describe experimental studies showing that sequential lineups are not more accurate than simultaneous lineups (Carlson & Carlson, 2014; Gronlund, Carlson, et al., 2013; Mickes et al., 2012; Dobolyi & Dodson, 2013). As mentioned, the WITNESS model is an example of a broader class of global matching models, which are motivated by Signal Detection Theory (SDT). The analyses used in recent simultaneous-sequential studies have a basis in models like WITNESS and the broader class of SDT models. Thus, a brief description of SDT follows.

Signal Detection Theory. SDT is a well-established theory that describes decision making under conditions of uncertainty (Macmillan & Creelman, 2005; Wickens, 2001). For example, in eyewitness identification, a witness must decide if any of the lineup members is the robber, despite issues like poor viewing conditions during the crime or a long time interval between the crime and the lineup. In a standard recognition memory task, participants study a set of targets and take a test that intermixes targets and distractors (never-before viewed items). Typically, one target or distractor is presented on each test trial. Participants can make one of two possible responses on a given trial, Old (identifying a lineup member) or New (rejecting the lineup). The calculations of identification and rejection rates are comparable to those illustrated for eyewitness identification experiments in Table 1. Each identification rate alone (correct identifications, false identifications, rejections) is an incomplete summary of performance. What is needed is an index that specifies discriminabilty—the ability to discriminate the presence versus the absence of the target. Such an index exists, and will be discussed shortly.

The degree of match to memory produced by a given target or distractor ranges from weak to strong. The variation in the degree of match to memory can be represented on a decision axis as depicted in the leftmost plot in Figure 1. The decision axis (labeled strength of evidence) spans higher degrees of match to memory on the right to lower degrees of match to memory on the left. Targets and distractors are drawn from separate distributions. The target distribution (in the context of eyewitness identification, the guilty distribution) is further to the right on the decision axis, indicating that items from the target distribution have higher match values on average compared to items from the distractor distribution (the innocent distribution). The target and distractor distributions overlap; thus, recognition memory test items drawn from the area of overlap will be difficult to discriminate. For example, the target and distractor items may look similar, or the participant may have attended less to some study items than others. The possible responses listed in Table 1 correspond to sections of the target and distractor distributions depicted in Figure 1.

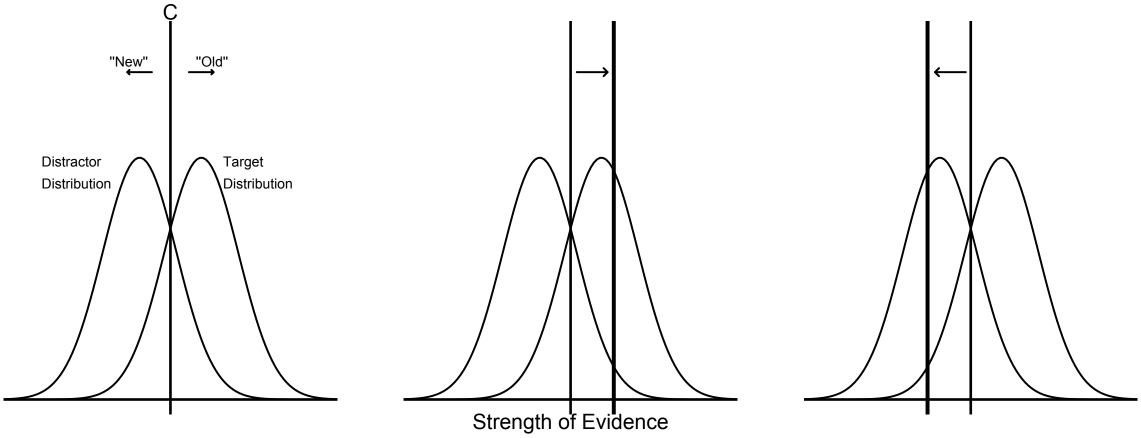


Figure 1. Example of SDT applied to recognition performance. The figure illustrates the theoretical shift in the placement of the criterion. The leftmost plot depicts optimal responding and the middle and rightmost plots show a conservative and liberal shift in criterion, respectively.

SDT represents discriminability based on the overlap between the target and distractor distributions. In contrast, SDT represents response bias based on the location of the decision criterion, the standard that specifies the level of match at which people judge an item to be Old. A criterion shift is a change in a person's willingness to make an Old judgment. People set and presumably shift their decision criterion depending on the degree to which environmental or experimental factors increase or decrease the amount of evidence required to accept an item as Old. As shown in the middle plot in Figure 1, individuals might adopt a higher, or conservative criterion, in which case the criterion would be set further to the right on the decision axis and only stimuli with higher familiarity values would be called Old (producing fewer correct and false identifications). Adopting a liberal, leftward criterion (rightmost plot of Figure 1) allows stimuli with lower familiarity values to be called Old (producing more correct and false identifications). SDT allows for an estimation of the placement of the criterion, C, calculated as $C = -\frac{1}{2}\left[z(\text{hit rate}) + z(\text{false alarm rate})\right]$. In eyewitness identification, higher values of C indicate that a lineup procedure shifts the criterion, yielding fewer identification responses. Lower values of C indicate that a lineup procedure yields more identification responses.

Measures of Accuracy

Two measures of discriminability (accuracy) derive from SDT: An index of the amount of overlap between the target and distractor distributions, d', and Receiver Operating Characteristic (ROC) analyses. The index d' is calculated by taking the difference of the z-scored correct and false identification rates, $d' = z(\text{correct ID}) - z(\text{false ID})$. d' measures the distance between the means of the guilty and innocent suspect distributions relative to the variance. To the extent that the signal and noise distributions overlap, d' will be lower indicating poorer discriminability, and complete overlap is signaled by $d' = 0$. The d' values for different lineup conditions can be compared to determine which procedure yields higher accuracy (Mickes, Moreland, Clark, & Wixted, 2014).

The best way to fully characterize eyewitness identification performance is to use receiver operating characteristic (ROC) curves (Mickes et al., 2012). ROC curves are plots of correct (y axis) and false identification rate (x axis) pairs at different eyewitness confidence cutoffs. Confidence cutoffs are used because confidence is assumed to be a proxy for memory strength, with higher confidence corresponding to higher memory strength. After making an identification response, witnesses in experiments estimate their confidence in their response on a scale (e.g. a scale from 1 = Not at all Confident to 6 = Very Confident).

The ROC curve is illustrated in Figure 2. The first point on an ROC (leftmost side of Figure 2) includes only correct and false identifications made at the highest level of confidence (6), the second point includes only correct and false identifications made at the highest and second highest level of confidence (6 and 5), and so on.

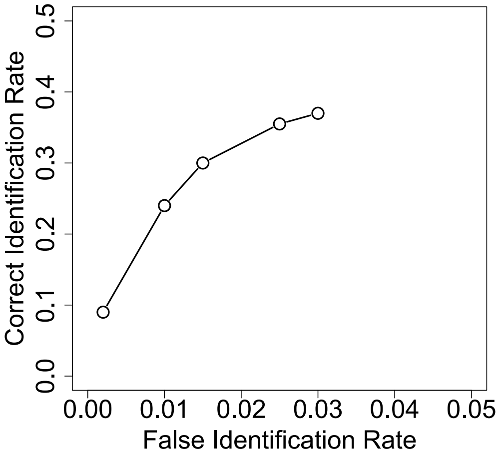


Figure 2. Example of an ROC plot for eyewitness identification.

Diagnostic accuracy is indicated by the area under the ROC curve. An ROC that tends toward the top left of the plot reflects better accuracy, or ability to discriminate guilty from innocent suspects, maximizing the correct identification rate and minimizing the false identification rate. Suspect ROC curves do not extend to the point where correct and false identification rates are both 1.0 because even if all witnesses identify someone from the lineup, some of those identifications will be identifications of foils. Thus, diagnostic accuracy is measured as the partial area under the curve, pAUC. pAUC analyses provide a statistical test of the degree to which the ROC curve for one lineup condition is higher than the ROC curve for another lineup condition.

Examining the family of correct and false identification rate pairs at each level of confidence provides a full understanding of eyewitness identification accuracy yielded by different lineup procedures. To the extent that the assumptions of SDT hold, d' approximates pAUC using a single correct and false identification rate pair. Earlier eyewitness identification studies such as Lindsay and Wells' (1985) simultaneous-sequential comparison relied on a measure that used a single correct and false identification rate pair, the ratio of correct to false identifications (C/F ratio), as a measure of accuracy. Lindsay and Wells computed the C/F ratio for sequential (3.06) and simultaneous lineups (1.35) and concluded that sequential lineups yielded higher accuracy. This is problematic because unlike SDT analyses, the C/F ratio does not distinguish between response bias and accuracy (Wixted & Mickes, 2014). The C/F ratio increases to the extent that the denominator is small. Thus, as the false identification rate tends towards zero, the C/F ratio increases irrespective of any proportional decrease of the correct identification rate. The results from meta-analyses (Clark, 2012; Palmer & Brewer, 2012; Steblay et al., 2011) show that sequential lineups reduce both the false and the correct identification rates. A procedure that decreases both the correct and false identification rates, like sequential lineup presentation, will yield a higher C/F ratio not because sequential presentation is more accurate than simultaneous presentation, but because sequential presentation produces a conservative response bias.

Application of SDT Analyses to Simultaneous-Sequential Comparisons

For 30 years, the dominant view has been that simultaneous lineups decrease accuracy, a view supported by studies that yielded lower C/F ratios for simultaneous lineups compared to sequential lineups (Steblay et al., 2001, 2011).

In addition to meta-analyses showing that sequential lineups reduce correct and false identification rates (Clark, 2012; Palmer & Brewer, 2012; Steblay et al., 2011), there is evidence that sequential superiority has declined over time. Clark, Moreland, and Gronlund (2013a) examined patterns of change in effect sizes of simultaneous and sequential comparisons across studies spanning the 29-year period beginning with Lindsay and Wells' (1985) seminal study. They computed cumulative effect size functions, which average over effect sizes from studies published up to and including a given year for all simultaneous-sequential comparisons. Several measures of effect size were considered, including d'. They found that the sequential advantage has largely disappeared over time since Lindsay and Wells' first study. Early studies showed a large accuracy advantage for sequential lineups, but recent studies showed either no advantage or a disadvantage. The decline in effect sizes for sequential lineups was attributed to a number of factors, including statistical issues, publication biases, and use of the C/F ratio.

Recently, Mickes et al. (2012) compared simultaneous and sequential lineups in three experiments and found a significant advantage of simultaneous lineups in one experiment and a nonsignificant advantage of simultaneous lineups in the other two experiments using ROC analyses. Since 2012, several empirical studies comparing simultaneous and sequential lineups showed no advantage for sequential lineups, and often showed an advantage for simultaneous lineups (Carlson & Carlson, 2014; Gronlund, Carlson, et al., 2013; Dobolyi & Dodson, 2013) based on ROC analyses. When simultaneous-sequential comparisons are analyzed using the C/F ratio, sequential lineups appear to be the better procedure. When analyzed using SDT-based analyses, which separate discriminability from bias, simultaneous lineups are as accurate or more accurate compared to sequential lineups.

What does the recent conclusion, that sequential lineups are not more accurate than simultaneous lineups, mean for the absolute-relative distinction? Sequential lineups were designed to prevent witnesses from making comparisons between lineup members, which in turn, should prevent witnesses from making relative judgments and therefore increase accuracy. However, sequential lineups do not increase accuracy. Is this because sequential lineups do not prevent relative judgments? Or, do relative judgments increase accuracy, rather than decreasing accuracy?

The Present Experiments

There are three issues with the absolute-relative theory of eyewitness decision-making that are addressed in this dissertation. First, much of the recent debate in eyewitness identification research concerning the positive or negative effects of absolute versus relative judgments has been argued with data from simultaneous and sequential lineups. However, it is important that theoretical inquiries do not become tied to a single experimental paradigm. Second, this paradigm yields conflicting findings, sequential superiority in some cases, and sequential inferiority in other cases. The finding of sequential inferiority is at odds with Wells' (1984) absolute-relative distinction, which assumes that comparisons among lineup members allowed by simultaneous lineups should decrease accuracy.

Third, the mapping between lineup presentation and decision rule is unclear. According to Lindsay and Wells (1985), sequential lineups should preclude comparisons among lineup members, thus precluding use of relative decision rules. However, people might learn which facial features are relevant and irrelevant over the course of the lineup, which should not be possible using a pure absolute decision rule. This is an issue because simultaneous-sequential studies may not adequately test the degree to which absolute judgments improve accuracy if sequential lineups do not actually prevent relative judgments. For example, Carlson, Gronlund, and Clark (2008) and Gronlund, Carlson, Dailey, and Goodsell (2009) found that identification performance was more accurate when the suspect appeared in late lineup positions in sequential lineups. It is possible that witnesses begin using “relative comparisons" (Gronlund, Andersen, & Perry, 2013) as the lineup progresses, or as witnesses gain experience with the facial features of the lineup members, they dynamically build a representation of each feature that improves discriminability in the later positions. Both explanations suggest that witnesses are able to make comparisons among lineup members and may even use relative decision rules in sequential lineups.

The present experiments were designed to examine the absolute-relative distinction and underlying comparative processes using two novel experimental paradigms. Two main questions were addressed: What effect do absolute and relative judgments have on performance? How do comparisons between lineup members influence identification accuracy and confidence?

In Experiment 1, participant-witnesses were presented with instructions to use an absolute or relative decision rule when making their identification decision. This simple paradigm empirically tests WITNESS model predictions about absolute and relative judgments. Experiment 2 investigated the comparative processes hypothesized to underlie relative judgments with a paradigm designed to influence the comparison process through the positioning of lineup foils relative to the suspect. Experiment 3 applied the foil-positioning paradigm to a comparison of simultaneous and sequential lineups. The specific aims and predictions for each experiment are detailed in the following chapters.

All experiments in this dissertation used experimental methods. Participant witnesses were presented with a staged crime video, followed by a distractor task and a subsequent lineup. Experiment 1 is presented in Chapter 2, Experiment 2 is presented in Chapter 3, and Experiment 3 is presented in Chapter 4. Chapter 5 presents a general discussion of the findings in the context of the eyewitness identification literature.

Chapter 2: Experiment 1

Relative judgments have long been assumed to reduce identification accuracy and increase false identification errors (Wells, 1984; Lindsay & Wells, 1985). In contrast, Wells predicted that absolute judgments would produce higher accuracy. The finding that a lineup procedure designed to prohibit relative judgments, sequential lineups, increased the C/F ratio compared to simultaneous lineups, has become the primary evidence for the absolute-relative distinction. As mentioned, the absolute-relative distinction is widely regarded as a theoretical cornerstone of eyewitness identification decision making research.

However, the absolute-relative distinction is underspecified and imprecise. It is unclear whether Wells' (1984) prediction that relative judgments decrease accuracy arises from any theory of memory. Clark et al. (2011) addressed these issues by generating identification response probabilities for absolute and relative decision rules for different foil selection strategies using the WITNESS model.

In actual criminal investigations, police choose foils that match the appearance of the suspect, termed a different-foils selection strategy in research. This strategy yields different foils for different suspects—namely, different foils depending on whether the suspect is the perpetrator or an innocent person. In the laboratory, the majority of eyewitness identification researchers use a same-foils selection strategy—choosing foils for the perpetrator and using those foils for the innocent suspect. The same-foils strategy yields the same foils for the perpetrator and an innocent suspect—in other words, the innocent suspect's lineup comprises foils that match the perpetrator. Traditionally, the same-foils design has been used to control for differences in lineup composition. In this manner, if one foil tends to draw identification responses (for example), that foil should affect responses similarly in both guilty and innocent suspect lineups. Clark and Tunnicliff (2001) have argued that the same-foils design is not the proper experimental control for equating foils across guilty and innocent-suspect conditions because the same-foils design does not simulate actual lineup administration.

To illustrate, assume that Steve Clark robs a liquor store, but the police suspect Curt Burgess (both are relatively tall, professorial white males). The same-foils design, applied in this situation, would produce a lineup with Curt Burgess as the suspect, and five foils who look like Steve Clark. How can it happen that the police would arrest the wrong person but select lineup foils who are similar to the right person? (It should almost never happen.) In real criminal investigations the foils are generally selected based on their similarity to the suspect; thus to the extent that the innocent suspect is not a clone of the perpetrator, lineups will contain different foils if the suspect is guilty than if the suspect is innocent. Moreover, in real criminal investigations, the foils should be as similar to the guilty suspect (when the suspect is guilty) as the foils are to the innocent suspect (when the suspect is innocent). This hair-splitting methodological difference produces a moderate difference in results. Clark and Tunnicliff (2001) showed that false identification rates were higher when foils were selected based on their match to the innocent suspect.

By formalizing absolute and relative judgments via the WITNESS model, Clark et al. (2011) provided clear definitions of absolute and relative processes in terms of decision rules. A witness using an absolute judgment should make an identification based only on the match to memory relative to a criterion, and a witness using a relative judgment should make an identification if the best match is better than the second best match, relative to a difference criterion. Formally, an absolute judgment is achieved by putting all of the decision weight on the best match and none on the difference between the best match and the next best match, $1.0(\text{BEST}) + 0.0(\text{BEST} - \text{NEXT}) < C$. A relative judgment is achieved by putting all of the decision weight on the difference between the best match and the next best match and none on the best match, $0.0(\text{BEST}) + 1.0(\text{BEST} - \text{NEXT}) < C_{\text{diff}}$. In both cases, an identification is made if the value exceeds the criterion.

Clark et al. (2011) modeled identification performance using a range of different parameter values that represented higher or lower accuracy of memory (e.g. $a = .3$ versus $a = .6$), and higher and lower similarity between the innocent suspect and the perpetrator and the foils and the perpetrator (e.g. $s[I,P]$ and $s[F,P]$ of .4 or .8). The similarity between the foils and the perpetrator $s[F,P]$ is the same between guilty and innocent-suspect lineups for same-foils lineups, because the foils are the same. The $s[F,P]$ parameter differs between guilty and innocent-suspect lineups for different-foils lineups, because the foils are less similar to the perpetrator when they are matched to the innocent suspect. Baseline parameters of $a = .3$, and $s[I,P]$ and $s[F,P]$ = .6 produce ROC curves that include correct and false identification rates from an experiment by Clark, Howell, and Davey

(2008), but in general, the mapping between parameter values and identification performance is not straightforward.

Clark et al. (2011) showed that the simulation for different-foils lineups produced an accuracy advantage for an absolute decision rule model over a relative decision rule model. The accuracy advantage of the absolute decision rule model depended on the foil similarity parameters. The foils in the different-foils innocent-suspect lineups are less similar to the perpetrator compared to the foils in guilty suspect lineups and same-foils innocent-suspect lineups (which are matched directly to the appearance of the perpetrator). According to the WITNESS model, lower foil-to-perpetrator similarity increases the false identification rate for the relative decision rule model. The next best match after the suspect is fairly dissimilar to the perpetrator, and is therefore a poor competitor with the innocent suspect.

Unlike the results for different-foils lineups, same-foils lineups produced inconsistent outcomes. Same-foils lineups produced an advantage of absolute decision rules when a was lower (.2) and foil similarity was high, ($s[F,P]=.8$). High foil-to-perpetrator similarity decreases the difference between the best match and the next best match to memory, decreasing performance of the relative judgment model. In contrast, same-foils lineups produced an advantage of relative decision rules when a was higher (.6) and the innocent suspect was similar to the perpetrator, ($s[I,P]=.8$). This may occur because a highly similar innocent suspect will yield a high best match value—and the absolute decision rule specifies to make an identification if the best match exceeds a criterion, increasing false identifications.

When lineups are constructed realistically using the different-foils design, the simulation results are consistent with Wells' (1984) hypothesis that absolute judgments improve identification accuracy. When modeling the same-foils design, the effect of decision rules on identification accuracy is less clear. This raises the concern that research using the same-foils design (the majority of research on eyewitness identification) might lead to different conclusions than research using the different-foils design.

Clark et al.'s (2011) simulations produced predictions about the effect of absolute and relative decision rules on identification behavior – and these predictions have never been tested directly. Experiment 1 in this dissertation provides an empirical test of the WITNESS model's predictions. Participant-witnesses were shown a staged crime, followed by a guilty-suspect, different-foils innocent-suspect, or same-foils innocent-suspect lineup. Witnesses first determined which of the lineup members was the best match to their memory of the perpetrator. After making this determination, witnesses were given instructions to use an absolute or relative decision rule to make their identification decision.

Predictions

If absolute judgments produce an advantage in different-foils lineups, as predicted by Clark et al.'s (2011) simulations, then witnesses who receive absolute decision rule instructions should be more accurate compared to those who receive relative decision rule instructions when the foils in the innocent-suspect lineup are selected based on their match to the innocent-suspect. For same-foils lineups, WITNESS predicted an advantage for an absolute decision rule model when memory was less accurate ($a = .2$) and similarity between the foils and the perpetrator was higher, ($s[F,P] = .8$). In contrast, WITNESS predicted an advantage for a relative judgment model when memory was more accurate, ($a = .6$) and the similarity between the innocent suspect and the perpetrator was higher, ($s[I,P] = .8$).

Method

Participants

Three thousand one hundred participants recruited from Amazon Mechanical Turk (AMT) participated for a remuneration of $3.00. Analyses are based on data from 2,956 participants. Of the original sample, 144 (4.6%) observations were excluded for the following reasons: 73 for technical problems such as buffering issues, computer crashes, or any other technical problem that prevented the participant from viewing the staged crime video, 32 for previously participating in a pilot study assessing the similarity relations in the lineups used in the study, 18 for not providing a description of the crime or perpetrator's appearance, 8 for providing a sufficiently vague description of the crime and perpetrator's appearance that indicated that they may not have seen the staged crime video, 7 for stating that the perpetrator was black (the perpetrator self identifies as Hispanic, though many participants identify his ethnicity as Caucasian), and 6 for participating twice (only data from their second attempt were excluded). Participants included in analyses were 30.6 (SD = 10.2) years old on average and 45.9% were female, 54.0% were male, and 0.2% self identified as transgendered or provided an unclear response. Participants self-identified as the following ethnicities: 75.6% Caucasian, 8.4% Asian/Pacific Islander, 7.3% African American, 5.3% Hispanic, 2.7% Mixed or biracial, 0.5% Native American, and 0.2% declined to state or provided an unclear response.

Materials and Procedure

Participants were asked to complete the study in a quiet room and to minimize distraction. There were no limitations on type of computer or browser; thus, participants used any electronic medium at their disposal. After signing up on AMT's webpage, participants were routed to a brief demographic questionnaire that queried their age, ethnicity, and whether their vision was corrected to 20/20. Participants then clicked a link that opened the experiment on a separate webpage. A video of a staged robbery at an ATM machine began to play automatically. The video was filmed from the perspective of a witness walking past a woman at an ATM. Shortly after passing the ATM, a male loud voice is heard off camera. The camera then turns back toward the noise, where the perpetrator is pointing a gun at the victim and demanding the money she had just withdrawn. She complies, dropping the money and her wallet on the ground. The perpetrator scrambles to retrieve the items and then races around a corner and out of the view of the camera. The video lasts 41s, with the perpetrator in full view for approximately 5s.

After the video, participants responded to the following open-ended questions: (1) Please type a description of the crime that you just witnessed, and (2) Please type a description of the robber's appearance (describe what the robber looked like). They were then queried regarding their confidence that they could recognize the robber if they saw him again (prospective confidence) on a 1 (not at all confident) to 6 (completely confident) scale. Participants then played Tetris during a five-minute delay interval prior to the lineup.

Lineup materials. Lineups were created with photographs obtained from the San Bernardino County California Sheriff's Department. To create a consistent appearance and minimize idiosyncratic details across lineup members, backgrounds were digitally altered to the same shade of blue, and clothing was digitally altered to show each lineup member in a black t-shirt. An innocent suspect was selected to have a high similarity to the guilty suspect. A separate group of 194 participants rated the similarity of 84 mugshots to either the guilty or innocent suspect. The lineups were constructed based on the ratings, as described below.

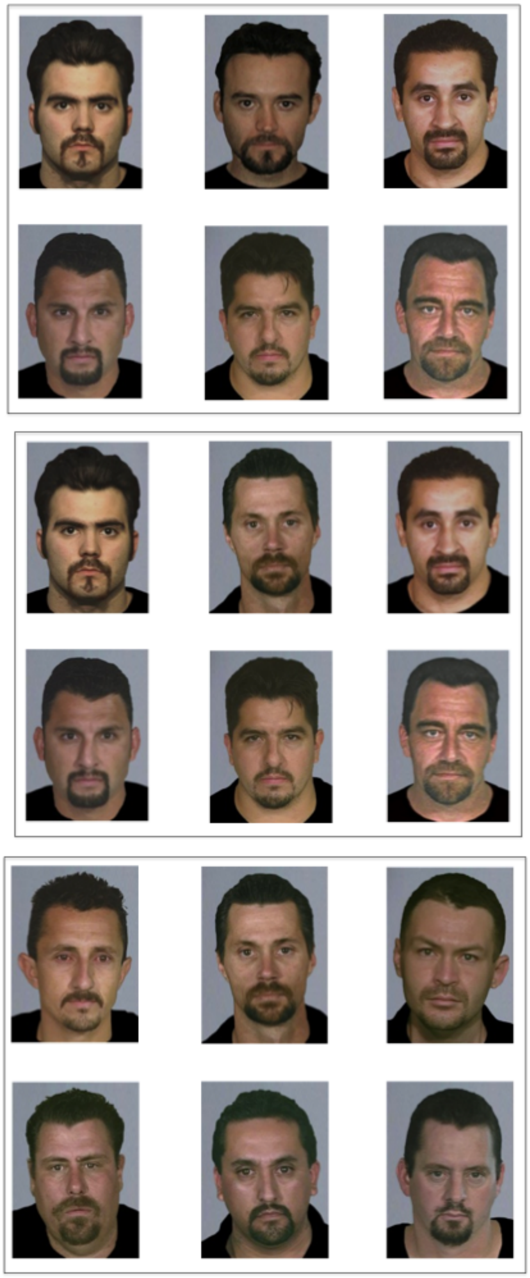


Figure 3. Lineups in Experiment 1. Guilty-suspect lineup appears on top, same-foils innocent-suspect lineup appears in the center, and different-foils innocent-suspect lineup appears on bottom. The suspect appears in position 2 in each.

Lineup construction. Examples of the lineups are shown in Figure 3. Each lineup contained a guilty or an innocent suspect and five foils. The critical compositional variable was foil selection strategy. Guilty-suspect lineups and different-foils innocent-suspect lineups used foils matched to the suspect's appearance (equated in terms of foil-to-suspect similarity). In contrast, same-foils innocent suspect lineups swapped out the guilty for the innocent suspect; thus, the foils were the same across the guilty suspect lineup and the same-foils innocent-suspect lineups. Foil-to-suspect similarity per lineup type is displayed in Table 2. The equated foil-to-suspect similarity between guilty lineups and different-foils innocent-suspect lineups is indicated by the Guilty-Suspect Different-Foils (To G_s) column, and the Innocent-Suspect Different-Foils (To I_s) column.

All lineups were presented in a two-by-three array, with positions 1-3 occupying the top row and 4-6 occupying the bottom row. The suspect appeared equally often in positions 1-6. Foil position was fixed with some exceptions. A foil that generally occupied a given position was shifted to the next position over depending on the position of the suspect. The positions of foils relative to the suspect's position are represented in Table 3. Foils were originally allocated to certain positions to maintain approximately equal foil-to-suspect similarity between rows. To illustrate, when the suspect was in position 1, lineup member 1 (the most similar) was in position 6, lineup member 2 was in position 2, lineup member 3 was in position 4, and so on.

Table 2

Foil-to-Suspect similarity per lineup type in Experiment 1. Each column sub heading indicates whether participants were rating mugshots to the mugshot of the guilty suspect (To G_s) or to the mugshot of the innocent suspect (To I_s).

	Guilty-Suspect Different-Foils	Innocent-Suspect Different-Foils		Innocent-Suspect Same-Foils
Rate to:	To G_s	To G_s	To I_s	To I_s
Suspect		4.36	3.60	3.60
Lineup Member 1	3.11	2.14	3.10	3.36
Lineup Member 2	2.98	2.92	2.96	2.12
Lineup Member 3	2.77	1.99	2.77	2.30
Lineup Member 4	2.64	2.41	2.60	2.22
Lineup Member 5	2.53	2.35	2.48	2.05

Table 3

Diagram of suspect and foil positions in Experiment 1. Lineup members (foils) are numbered in terms of similarity such that lineup member 1 is the most similar member to the suspect and lineup member 5 is the least similar lineup member to the suspect.

Suspect Position	Lineup Member 1	Lineup Member 2	Lineup Member 3	Lineup Member 4	Lineup Member 5
1	6	2	4	3	5
2	6	1	4	3	5
3	6	1	4	2	5
4	6	1	3	2	5
5	6	1	3	2	4
6	5	1	3	2	4

Lineup procedure. The lineup instructions were presented on the screen and read over headphones by a recorded voice. Participants were told to look at all photographs before responding and that the person who committed the crime may or may not appear in the lineup. After pressing a button to advance the screen, the lineup appeared on the screen with additional instructions that asked, “Of the six people in the lineup, which person is the best match to your memory of the perpetrator?" Participants responded by typing the number that appeared above a given lineup member’s photograph. After making the best-match determination, participants heard and saw either Absolute or Relative decision instructions. The process that computes the match between each lineup member and memory is separate from the decision rules in the WITNESS model; thus, in the present study, the best-match determination occurred prior to the decision rule manipulation. Both instruction types were presented in two syntactically similar sentences, the first indicating the rule to identify and the second indicating the rule to reject the lineup. For example, the Absolute decision instructions directed participants to identify the best match if “that [best-matching] lineup member is a sufficiently good match to your memory of the perpetrator." Relative decision instructions directed participants to identify the best match if that lineup member was “a sufficiently better match to your memory of the perpetrator than anyone else in the lineup." In the relative condition, all lineup members remained on screen; whereas in the absolute condition, only the lineup member identified as the best match remained on screen during the identification decision. Participants responded by clicking a radio button located above the lineup member to make an identification, or clicked a button that said “the person who committed the crime is not in the lineup."

Participants then rated their retrospective confidence in their response on a 1(not at all confident) to 6 (completely confident) scale. Participants pressed a key to indicate their responses to identification questions and to rate retrospective confidence. After the identification procedure, the lineup was displayed again and participants rated the similarity of each lineup member to their memory of the robber in the video using a scale of 1 (lineup member looks very little like your memory of the robber's appearance) to 100 (lineup member looks very much like your memory of the robber's appearance) scale. Participants typed a number between 1 and 100 for each lineup member.

Results

The following section presents three sets of analyses. The first set examined differences in discriminability by instruction condition via d' and ROC analyses (ROC curves and pAUC comparisons). This set of analyses directly addressed the central question of this research: How do decision rules affect eyewitness accuracy? The second set of analyses compared identification performance of witnesses who received absolute and relative instruction conditions. These log-linear analyses assessed the extent to which the instructions to use a given decision rule affected suspect identifications, foil identifications, and non-identification rates for guilty and same-foils and different-foils innocent-suspect lineups. The third set of analyses investigated eyewitness confidence by analyzing the relationship between confidence and accuracy.

Discriminability

This set of analyses evaluated the effect of absolute versus relative decision rules on identification accuracy using d', ROC curves, and pAUC. Using the false identification rate from different-foils lineups, the d' for relative decision rules (0.497) was slightly but not significantly higher than the d' for absolute decision rules (0.346), $G = 1.294$, $p = .196$, using the method defined by Gourevitch and Galanter (1967). Similarly, using the false identification rate from same-foils lineups, the d' for relative decision rules (0.541) was slightly but not significantly higher than the d' for absolute decision rules (0.405), $G = 1.161$, $p = .246$. Aggregating the false identification

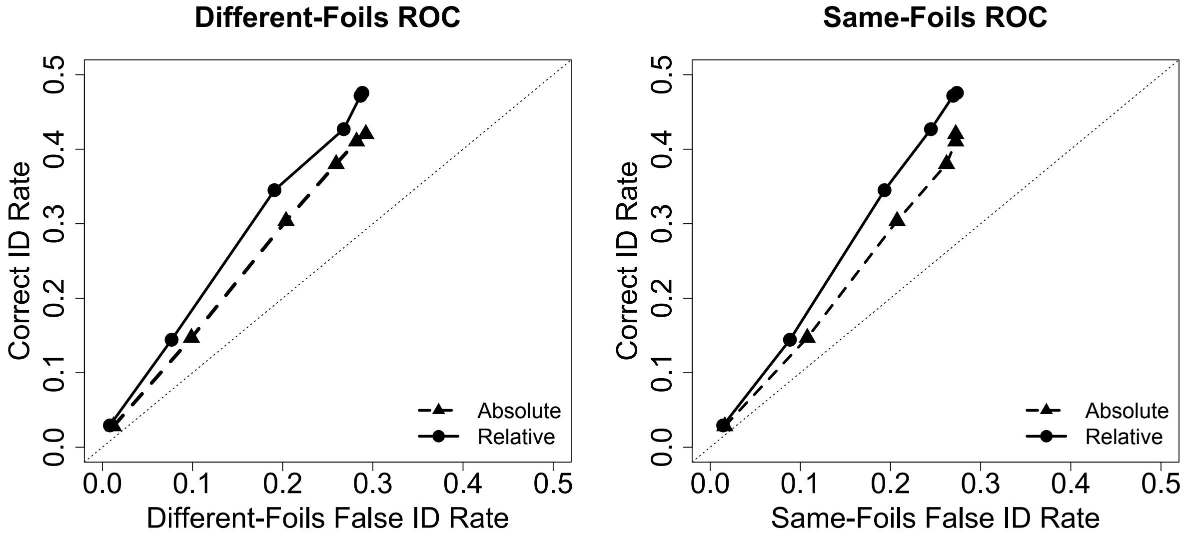


Figure 4. ROC plots for Different-Foils lineups (left) and Same-Foils lineups (right). Absolute decision rules are indicated by solid lines, relative decision rules are indicated by dashed lines.

rates for same- and different-foils designs yielded a nonsignificant accuracy advantage for relative decision rules (0.519) over absolute decision rules (0.376), $G = 1.436$, $p = .151$.

ROC curves for absolute and relative instruction conditions and different-foils and same-foils lineups are shown in Figure 4. The curves had a linear appearance, which can be indicative that the underlying assumptions of SDT were not met. The WITNESS model, however, produces a linear ROC under certain circumstances. For example, in Clark et al.'s (2011) simulations, when memory accuracy was low, $s[I,P]$ was higher and $s[F,P]$ was lower, same-foils lineups were predicted to yield linear ROCs for both absolute and relative decision rules. Similarly, for different-foils lineups, the predicted ROCs had a linear appearance when memory accuracy was low, $s[I,P]$ was higher and $s[F,P]$ was moderate to high.

The pAUC for instructions and lineup conditions were compared using the pROC program developed by Robin et al. (2011). For different-foils lineups, the pAUC was nonsignificantly larger for relative decision rules (.124) than for absolute decision rules (.108), $D = 1.312$, $p = .189$. For same-foils lineups, the pAUC was nonsignificantly larger for relative decision rules (.124) than for absolute decision rules (.107), $D = 1.323$, $p = .186$. Combining same- and different-foils innocent suspect lineups yielded a nonsignificantly higher pAUC for relative decision rules (0.124) compared to absolute decision rules (0.108), $D = 1.595$, $p = .111$.

Table 4 displays d' and pAUC comparisons by suspect position, lineup type, and combined same- and different-foils lineups. The pattern of a small advantage of relative decision rules was consistent across most suspect positions, with the following exceptions. For different-foils lineups: d' was nonsignificantly lower for relative decision rules in suspect position 2 and pAUC was not, and for suspect position 4, both d' and pAUC were nonsignificantly lower for relative compared to absolute decision rules.

For both same-foils lineups and combined lineups: both d' and pAUC showed a small advantage for absolute decision rules in suspect position 2. For same-foils lineups, there was a small advantage for absolute decision rules in suspect position 5 for pAUC only. For both different-foils lineups and combined lineups, d' and pAUC showed a statistically significant advantage for relative decision rules in suspect position 6. To examine the overall trend of the position effects, a one-sample t-test compared the differences between absolute and relative decision rules for each lineup type, for both d' and pAUC. Only the test combining pAUC for the combined false identification rates was marginal, $t(5) = 2.538$, $p = .052$, $r = .750$.

The discriminability results indicated that there was a small advantage of relative decision rules—but that advantage was not consistent across suspect position. Further, suspect position influenced the relationship between foil selection strategy (foil-to-suspect similarity) and accuracy. A discussion of the inconsistency across suspect position appears in the discussion section.

Identification Analyses

The identification analyses examined patterns of associations between identification responses, instruction conditions, lineup conditions, and suspect position. These analyses can be used in conjunction with discriminability analyses to clarify the degree to which differences in accuracy are due to changes in the correct and false identification rates (Carlson & Carlson, 2014). Proportions of identifications of suspects, foils, and non-identification responses for absolute and relative conditions appear in Table 5. Data were analyzed using separate log linear analyses for

Table 4

d' and pAUC and corresponding statistics across suspect positions for absolute and relative decision rules and same-foils, different-foils, and combined false identification conditions. Abs = absolute, Rel = Relative, Stat indicates the test statistic value (G for d' and D for pAUC), and p indicates the p-value for the comparison. Separate one-sample t-tests for both d' and pAUC meta-analytically aggregate test-statistics across suspect positions.

		Different-Foils				Same-Foils				Combined			
Position		Abs	Rel	Stat	*p*	Abs	Rel	Stat	*p*	Abs	Rel	Stat	*p*
1													
	d'	0.522	0.801	0.983	.326	0.609	0.773	0.568	.570	0.565	0.790	0.908	.364
	pAUC	0.138	0.159	0.651	.515	0.144	0.155	0.341	.733	0.122	0.157	1.384	.166
2													
	d'	0.547	0.515	-0.110	.912	0.604	0.446	−0.552	.581	0.575	0.480	−0.387	.699
	pAUC	0.127	0.129	0.068	.946	0.122	0.116	−0.201	.840	0.125	0.123	−0.081	.936
3													
	d'	0.303	0.407	0.372	.710	0.346	0.474	0.461	.645	0.324	0.441	0.492	.623
	pAUC	0.089	0.115	0.885	.376	0.091	0.116	0.857	.392	0.090	0.115	1.035	.301
4													
	d'	0.319	0.098	−0.782	.434	0.319	0.585	0.911	.362	0.319	0.328	0.035	.972
	pAUC	0.098	0.092	−0.206	.837	0.099	0.139	1.341	.180	0.098	0.113	0.590	.555
5													
	d'	0.361	0.587	0.784	.433	0.449	0.484	0.121	.903	0.406	0.534	0.528	.597
	pAUC	0.110	0.125	0.492	.622	0.111	0.102	−0.281	.779	0.111	0.114	0.118	.906
6													
	d'	0.032	0.652	2.105	.035	0.104	0.545	1.505	.132	0.068	0.597	2.095	.036
	pAUC	0.083	0.141	2.018	.043	0.076	0.125	1.680	.093	0.079	0.133	2.161	.031
d'													
one sample *t*				1.382	.226			1.763	.138			1.757	.139
pAUC													
one sample *t*				2.052	.094			1.889	.118			2.538	.052

Table 5

Proportions and frequencies of suspect, foil, and non-identification responses for absolute and relative conditions by suspect guilt and foil selection strategy in Experiment 1.

	Guilty-Suspect Different-Foils		Innocent-Suspect Different-Foils		Innocent-Suspect Same-Foils	
Response	Abs	Rel	Abs	Rel	Abs	Rel
Suspect	209 (.421)	244 (.476)	142 (.292)	139 (.288)	134 (.272)	133 (.274)
Foil	91 (.183)	100 (.195)	149 (.307)	131 (.272)	124 (.252)	132 (.272)
Reject	197 (.396)	169 (.329)	195 (.401)	212 (.440)	234 (.476)	221 (.455)

guilty-suspect, same-foils innocent-suspect, and different-foils innocent-suspect lineups with response type (suspect identifications, foil identifications, and no-identifications), instruction condition (absolute and relative) and suspect position (1-6) entered as predictors. The following reports likelihood ratio chi-square statistics for log linear analyses and Pearson chi square statistics for follow-up comparisons.

Guilty Suspect lineups.

Correct Identifications. The correct identification rate was slightly but nonsignificantly higher when witnesses were given a relative decision rule (.476) than when witnesses were given an absolute decision rule (.421), $\chi^2(1, N = 1,010) = 3.102$, $p = .078$, $r = .055$. The highest-order interaction in the log-linear model (identification response × instructions × suspect position) was not statistically reliable, indicating that the effect of instructions on response did not vary with the position of the suspect, $\chi^2(5, N = 1,010) = 6.330$, $p = .275$. No effects were statistically significant ($.113 < p < .630$).

Foil Identifications. The foil identification rate was slightly but nonsignificantly higher when witnesses were given a relative decision rule (.195) than when witnesses were given an absolute decision rule (.183), $\chi^2(1, N = 1,010) = 0.231$, $p = .631$, $r = .015$. The highest-order interaction between identification response, instructions, and suspect position was not statistically reliable, $\chi^2(5, N = 1,010) = 2.156$, $p = .827$. No effects were statistically significant ($.189 < p < .658$).

Lineup rejections. The lineup rejection rate was significantly lower when witnesses received a relative decision rule (.329) than when witnesses received an absolute decision rule (.396), $\chi^2(1, N = 1,010) = 4.899$, $p = .027$, $r = .070$, indicating that witnesses were less likely to incorrectly reject the guilty suspect lineup when using relative decision rules. The highest-order interaction was not statistically reliable, $\chi^2(5, N = 1,010) = 2.748$, $p = .739$. No other effects achieved statistical significance ($.671 < p < .696$).

Different-Foils Innocent-Suspect lineups.

False Identifications. The false identification rate was slightly but not significantly lower for relative decision rules (.288) compared to absolute decision rules (.292), $\chi^2(1, N = 968) = 0.017$, $p = .896$, $r = .004$. No effects reached statistical significance ($.531 < p < 1.00$).

Foil Identifications. The foil identification rate was nonsignificantly lower for relative decision rules (.272) than for absolute decision rules (.307), $\chi^2(1, N = 968) = 1.426$, $p = .232$, $r = .038$. No effects reached statistical significance ($.256 < p < 1.00$).

Lineup Rejections. Witnesses were slightly but not significantly more likely to correctly reject the different-foils innocent suspect lineup when using relative decision rules (.440) compared to absolute decision rules (.401), $\chi^2(1, N = 968) = 1.480$, $p = .224$, $r = .039$. No effects, including the highest order interaction, were statistically significant ($.337 < p < 1.00$).

Same-Foils Innocent-Suspect lineups.

False Identifications. Although there was one less false identification in the relative decision rule condition (133) compared to the absolute decision rule condition (134), the false identification rate was slightly higher, proportionally, when witnesses were given a relative decision rule (.274) than when witnesses were given an absolute decision rule (.272), $\chi^2(1, N = 978) = 0.002$, $p = .964$, $r = .001$. The discrepancy between the frequencies and proportions was an artifact of unbalanced cells—slightly fewer witnesses received relative instructions (.497) compared to absolute instructions (.503). No effects in the model reached statistical significance ($.339 < p < 1.00$).

Foil Identifications. The foil identification rate was slightly but nonsignificantly higher when witnesses were given a relative decision rule (.272) than when witnesses were given an absolute decision rule (.252), $\chi^2(1, N = 978) = 0.485$, $p = .486$, $r = .022$. No effects in this model reached statistical significance ($.781 < p < 1.00$).

Lineup Rejections. Witnesses were slightly but not significantly less likely to correctly reject the same-foils innocent suspect lineup when using relative decision rules (.455) than absolute decision rules (.476), $\chi^2(1, N = 978) = 0.428$, $p = .513$, $r = .021$. No effects, including the highest order interaction, were statistically significant ($.713 < p < 1.00$).

Response Change from Best-Match to Identification. Upon viewing the lineup, witnesses were asked to determine the best-match to memory, after which witnesses received absolute or relative instructions to identify the best match or reject the lineup. This section presents analyses of the proportion of witnesses who ultimately identified the best matching lineup member after receiving absolute or relative decision rules.

Guilty-Suspect Lineups. Of the 320 witnesses who identified the suspect as the best match in the absolute condition, 209 (.653) identified him as the perpetrator. Similarly, of the 344 witnesses who identified the suspect as the best match in the relative condition, 244 (.716) identified him as the perpetrator. A two-sample test for equality of proportions (a χ^2 comparison of two proportions) showed that witnesses who received relative instructions were marginally more likely to identify the guilty suspect compared to those who received absolute instructions, $\chi^2(1, N = 661) = 2.982$, $p = .084$, $r = 0.067$.

Innocent-Suspect Lineups. For different-foils lineups, of the 223 witnesses who identified the suspect as the best match in the absolute condition, 142 (.637) identified him as the perpetrator, and of the 231 witnesses who identified the suspect as the best match in the relative condition, 139 (.602) identified him as the perpetrator. Witnesses who received relative instructions were nonsignificantly less likely to incorrectly identify the innocent suspect as the perpetrator compared to absolute instructions, $\chi^2(1, N = 454) = 0.591$, $p = .442$, $r = .036$.

In the same-foils lineups, of the 227 witnesses who identified the suspect as the best match in the absolute condition, 134 (.590) identified the suspect as the perpetrator, and of the 232 witnesses who identified the suspect as the best match in the relative condition, 133 (.573) identified the suspect as the perpetrator. As with the different-foils lineups, witnesses who received relative instructions were nonsignificantly less likely to falsely identify the innocent suspect, $\chi^2(1, N = 459) = 0.137$, $p = .712$, $r = .017$.

Suspect Position. The proportions of witnesses who subsequently identified the suspect as the perpetrator after receiving absolute and relative instructions were compared using the test of two proportions within suspect position for each lineup type. The proportions and corresponding p-values appear in Table 6.

Table 6

Proportion of witnesses who identified the suspect after receiving absolute or relative instructions by lineup type and suspect position. p-value for comparison of absolute and relative proportions within suspect position. An asterisk follows the p-value when the proportions were numerically higher for absolute compared to relative decision rule conditions.

	Guilty-Suspect Different-Foils			Innocent-Suspect Different-Foils			Innocent-Suspect Same-Foils		
Position	Abs	Rel	p	Abs	Rel	p	Abs	Rel	p
1	0.639	.750	.179	0.629	.561	.550*	.541	.511	.790*
2	0.750	.774	.777	0.611	.610	.990*	.677	.628	.660*
3	0.704	.652	.544*	0.789	.568	.039*	.737	.645	.410*
4	0.680	.680	1.000*	0.686	.721	.734	.571	.462	.323*
5	0.614	.714	.259	0.526	.606	.499	.475	.590	.307
6	0.522	.731	.032	0.585	.528	.612*	.564	.629	.573

For guilty-suspect lineups, four of the six positions yielded a slight advantage for relative decision rules (positions 1, 2, 5, and 6) but the only statistically significant increase in the correct identification rate for relative decision rules occurred when the suspect appeared in position 6 ($p =$ .032). Both innocent-suspect lineup types displayed a slight nonsignificant advantage for relative decision rules across most suspect positions. Specifically, somewhat fewer witnesses who received relative instructions falsely identified the innocent suspect when he appeared positions 1, 2, 3 and 5 for different-foils lineups and positions 1-4 for same-foils lineups. Thus, the slight advantage of relative decision rules was generally consistent across four of six possible suspect positions for all lineup types.

Foil Selection Strategy. Clark and Tunnicliff (2001) found that the same-foils design used by eyewitness identification researchers underestimated the false identification rate compared to the different-foils design, which suggests that most research on eyewitness identification may yield artifactually low false identification rates. The present study included a comparison of the two foil selection strategies. False identifications were nonsignificantly higher using the different-foils design (.290) compared to the same-foils design (.273) collapsing over instructions and suspect position, $\chi^2(1, N = 1,946) = 0.718$, $p = .397$, $r = .019$. Neither comparison within lineup instructions nor suspect position yielded statistical significance ($.10 < p < .96$). Possible reasons for lack of statistical significance will be considered in the Discussion section.

Witnesses made fewer correct rejections for different-foils lineups (.420) compared to same-foils lineups (.465), $\chi^2(1, N = 1,946) = 3.954$, $p = .047$, $r = .045$. This effect depended on instructions: Correct rejections were significantly lower in different-foils lineups (.401) than same-foils lineups (.476) for absolute instructions $\chi^2(1, N = 978) = 5.492$, $p = .019$, $r = .075$, but not relative instructions (different-foils lineups: .455; same-foils lineups: .440), $\chi^2(1, N = 968) = 0.217$, $p = .641$, $r = .015$. The finding of fewer correct rejections and slightly more false identifications in different-foils lineups provided some support for the prediction that the false identification rate for same foils would be lower than the false identification rate for different foils. This prediction suggests that selecting foils to match the innocent suspect in an attempt to protect the suspect may backfire, because the foils that match the innocent suspect are less similar to the perpetrator than the innocent suspect. The lower similarity between the innocent suspect's foils and the perpetrator may increase the best match value more in the different-foils condition compared to the same-foils condition, decreasing the correct rejection rate for different foils in the absolute condition (which relies on the comparison between the best match and a criterion).

Confidence

Confidence and accuracy.

The analysis of witness confidence is important because confidence is used by trial courts to make decisions about whether identification evidence is sufficiently reliable to present to the jury, and because pattern jury instructions in many states instruct jurors that they may consider witness confidence in determining how much weight to give to the identification evidence. Further, the relationship between confidence and accuracy is an assumption of SDT and is foundational to ROC analyses. The relationship between confidence and accuracy was assessed by calculating point-biserial correlations and by plotting the calibration curves. The point-biserial correlations, considering only positive identifications (suspect and foil identifications) were $r = .038$ ($N = 777$), $p = .284$ for relative decision rules and $r = .021$ ($N = 756$), $p = .564$ for absolute decision rules. These correlations are not significantly different from each other, $z = .340$, $p = .734$. Considering only suspect identifications, the correlations were $r = .029$, $p = .510$ for relative decision rules and $r = -0.015$, $p = .750$ for absolute decision rules. The difference between the two correlations was not significant ($p = .490$). The weak relationships between confidence and accuracy for the different decision rules statistically may have obscured the relationship between confidence and accuracy because the distribution of responses varied across confidence levels (Juslin, Olsson, & Winman, 1996). Juslin et al. (1996) and more recently, Wixted, Mickes, Clark, Gronlund, and Roediger (In Press) have suggested that calibration curves provide a better measure of the relationship between confidence and accuracy.

The confidence-accuracy calibration curves are shown in Figure 5. In the figure, the vertical axis shows accuracy as the proportion of suspect identifications that were correct (correct / [correct + false]). The horizontal axis represents confidence on a 1-to-6 scale. Few witnesses who made

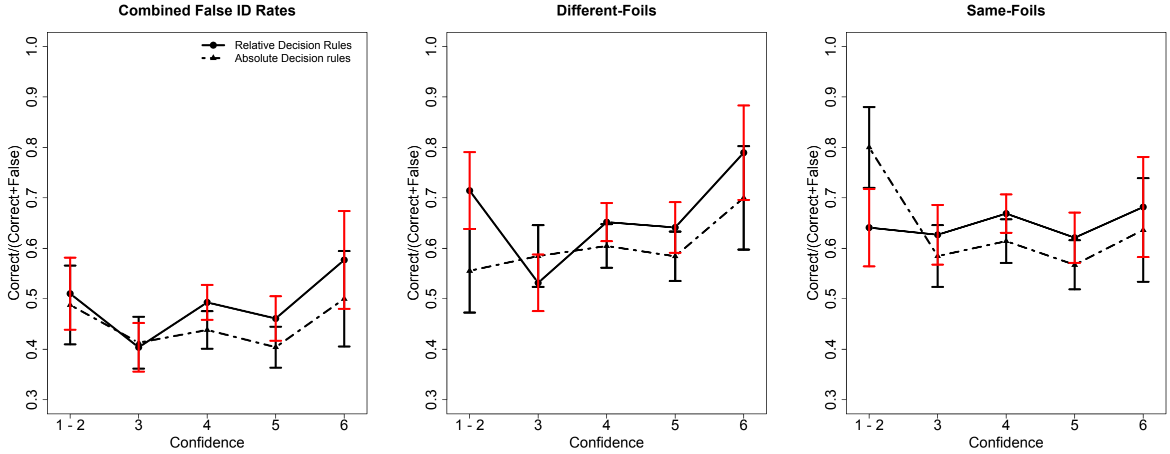


Figure 5. *Confidence-accuracy calibration curves for combined false identification rate, different-foils, and same-foils lineups. Error bars represent the standard error of a proportion. Absolute decision rule error bars appear in black, relative decision rule error bars appear in red.*

identification responses reported a confidence level of “1," and thus, confidence values of “1" were combined with “2" for all lineup types.

In addition to the confidence-accuracy calibration curves, correlations between confidence and accuracy, with accuracy defined as the proportion of suspect identifications that were correct, are reported as well. Although the slope of the confidence-accuracy calibration curve was not particularly steep, accuracy generally increased with confidence for different-foils lineups in the absolute condition, $r = .822$, $p = .088$, and relative condition, $r = .431$, $p = .469$. Relative decision rules produced an increase in accuracy at the lowest level of confidence. For same-foils lineups, the confidence-accuracy calibration curves were fairly flat for relative decision rules, $r = .453$, $p = .444$, and negative for absolute decision rules, $r = -0.586$, $p = .230$, because absolute decision rules showed an increase in accuracy at the lowest levels of confidence. The apparent interaction between decision rules and lineup type that showed higher accuracy at low levels of confidence for relative rules (different-foils lineups) and absolute rules (same-foils lineups) may have been due to low numbers of responses at confidence levels of 1 and 2. Figure 6 shows the distributions of responses per confidence level. For relative decision rules, there were only 10 low confidence false identifications for different-foils lineups, 14 for same-foils lineups, but 25 correct identifications. For absolute decision rules, there were 16 false identifications for different-foils lineups, 5 for same-foils lineups, and 20 correct identifications. A chi-square test comparing correct and false identifications made with confidence of 1 and 2 for absolute and relative instructions was not statistically significant for different-foils lineups, $\chi^2(1, N = 71) = 1.927$, $p = .165$, $r = .165$, or for same-foils lineups, $\chi^2(1, N = 64) = 1.845$, $p = .174$, $r = .170$. The lack of statistical significance for these tests suggested that the interaction between decision rule and lineup type may have been spurious.

For different-foils lineups, the confidence-accuracy results suggested that higher confidence tended to signal higher accuracy for both absolute and relative decision rules. The relationship was less clear for same-foils lineups, which produced flatter curves for both decision rules. For different-foils lineups, accuracy was slightly higher for those who made identification responses with a confidence of 6 for absolute (.789) and relative (.700) decision rules compared to same-foils lineups (absolute = .682, relative = .636). Because the foils were more similar to the perpetrator in the same-foils case, the innocent suspect had more competition; thus, false identifications should have been made with lower confidence. However, the false identifications were made with higher confidence in the same-foils condition and lower confidence in the difference foil condition, as depicted in Figure 6. In the present study, the innocent suspect was a much better match to the perpetrator than any of the foils based on similarity ratings. The mean rating of innocent suspect to the perpetrator was 4.36, whereas the best matching lineup member was rated as having a similarity of 3.11 to the perpetrator. The relatively low foil-to-perpetrator similarity may have had less of an impact than foil-to-suspect similarity. In this study, confidence may have been lower for false identifications made in the different-foils condition because the higher similarity between the innocent suspect and foils matched to his appearance may have helped witnesses ignore shared features, and focus on diagnostic features, increasing accuracy at higher levels of confidence.

Discussion

Experiment 1 investigated the effect of absolute and relative decision rules on identification accuracy and provided an empirical test of the WITNESS model predictions generated by Clark et al. (2011). In contrast to the prediction of an absolute judgment advantage, the identification results showed a small advantage of relative judgment instructions. Witnesses instructed to use a

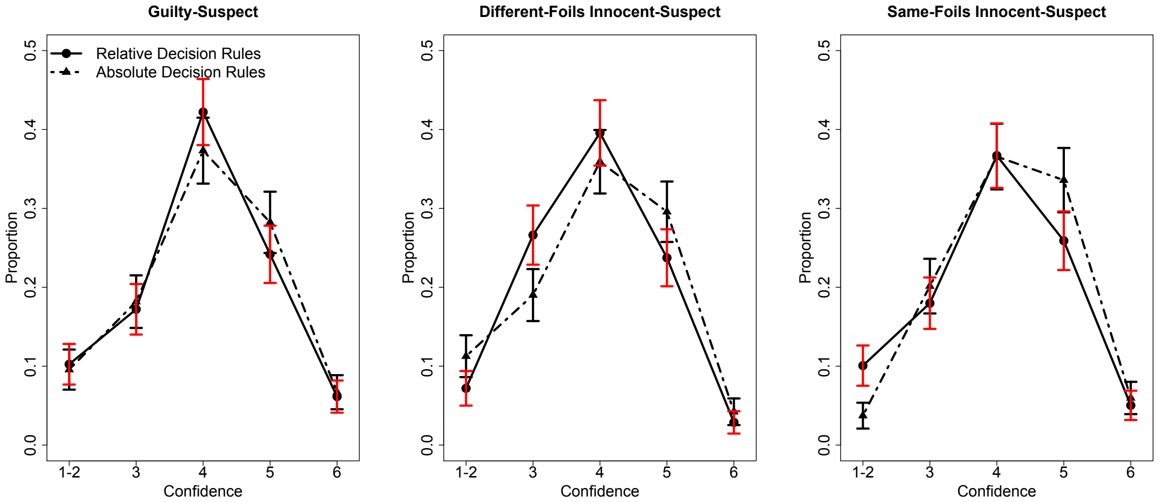


Figure 6. *Plots of proportions of suspect identification responses per confidence level by lineup type and decision rule. Error bars represent the standard error of a proportion. Absolute decision rule error bars appear in black, relative decision rule error bars appear in red.*

relative decision rule made marginally more correct identifications and fewer incorrect rejections of the guilty suspect compared to those instructed to use an absolute decision rule. Further, witnesses instructed to use a relative decision rule were slightly less likely to falsely identify the innocent suspect and slightly more likely to correctly reject the innocent suspect lineup in different-foils lineups compared to those instructed to use an absolute decision rule. The opposite pattern was found for same-foils lineups: Witnesses instructed to use a relative decision rule were slightly more likely to falsely identify the innocent suspect and slightly less likely to correctly reject the innocent suspect lineup compared to those instructed to use an absolute decision rule. When considering only witnesses who identified the innocent suspect as the best match prior to receiving the decision rule manipulation, slightly fewer witnesses who received relative instructions opted to identify the innocent suspect in both same- and different-foils lineups.

The discriminability results were consistent with the identification results, and showed a non-significant advantage for relative decision rules in terms of ROC curves, pAUC, and d'. This small advantage varied across suspect positions. For different-foils lineups, pAUC results showed a non-significant accuracy advantage for relative decision rules in suspect positions 1, 2, 3, and 5, and a statistically significant accuracy advantage in suspect position 6. A small accuracy disadvantage was found for relative decision rules in suspect position 4.

One possible explanation for the variation in the effect of decision rules across suspect position involves the suspect's neighborhood similarity. The suspect's neighborhood refers to the foils nearest the suspect. As will be discussed in detail, Experiment 2 showed that neighborhood similarity can impact identification accuracy—specifically, surrounding the suspect with high-similarity foils increased the accuracy of identification responses compared to surrounding the suspect with low-similarity foils. When high-similarity lineup members surrounded the suspect, witnesses may have compared these best matching lineup members to each other to determine the facial features that are most diagnostic of guilt, thus improving accuracy.

In Experiment 1, the positions of the foils depended on the position of the suspect, such that when the suspect shifted position, his neighbors (in this case, the foils adjacent, above or below the suspect) changed systematically. This may have created configurations that differed in neighborhood similarity. If neighborhood is defined as the foils that are easiest to compare–those nearest to the suspect–then when the suspect is in the central positions (2 and 5), the neighbors are to the left and right of the suspect. If the suspect is on the edge (1, 3, 4, and 6) then one nearest neighbor sits to the left or right, and another nearest neighbor is above or below. Of course, neighborhood could assume different definitions, but these definitions will be used for the sake of simplicity. To illustrate, the similarity of the suspect's two nearest neighbors was averaged for each of the six suspect positions, yielding the following overall similarity neighborhoods per suspect position in the guilty-suspect condition (the similarity relations between the guilty and different-foils innocent-suspect lineups were equated to be as close as possible): Position 1 = 2.88, Position 2 = 2.81, Position 3 = 2.88, Position 4 = 2.76, Position 5 = 2.82, Position 6 = 2.94. The lowest similarity configuration occurs when the suspect is in position 4—the foil in position 1 (above the suspect) has a similarity of 2.98 and the foil to the suspect's right has a similarity of 2.53. This, incidentally, was the position that yielded a pAUC accuracy disadvantage for relative judgments for different-foils lineups. In contrast, the highest similarity configuration of the neighbors occurred in suspect position 6—the foil in position 3 (above the suspect) had a rated similarity of 2.77 and the foil to the left of the suspect had a rated similarity of 3.11—which yielded a significant pAUC advantage for relative judgments for different-foils lineups. Same-foils lineups yielded slight accuracy disadvantages for relative decision rules in suspect positions 2 and 5, and a marginal advantage of relative decision rules in suspect position 6. The similarity relations between guilty and same-foils innocent-suspect lineups were different—the same-foils lineups contained lower similarity configurations in general, ranging from 2.09 (position 3) to 2.83 (position 6) that differed somewhat in rank ordering compared to different-foils lineups, obscuring the potential impact of neighborhood.

In Experiment 1, in different-foils lineups, the suspect's neighborhood may have moderated the influence of instructions on identification accuracy. When instructed to use relative judgments in suspect position 4, the comparisons among lineup members allowed by relative judgments between lower similarity lineup members may have been less diagnostic of guilt, decreasing accuracy. In contrast, when the comparisons among lineup members allowed by relative judgments were between the similar lineup members, as in suspect position 6, the comparisons among lineup members may have been more diagnostic of guilt, increasing accuracy. Although the possibility that neighborhood may interact with decision rules is intriguing, future studies comparing absolute and relative judgments should randomize the positions of the foils to control for potential confounds due to neighborhood. The influence of the neighbors will be considered more fully in Experiments 2 and 3.

Confidence-accuracy calibration results showed that accuracy was somewhat higher at all but the lowest levels of confidence for relative compared to absolute decision rules for both innocent-suspect lineup types. Confidence and accuracy were moderately correlated for both absolute and relative decision rules in different-foils lineups, indicating that confidence is a reasonable proxy for accuracy when foils are selected based on their similarity to the suspect. For same-foils lineups, the absolute and relative decision rule curves were both fairly flat, and negatively correlated in the absolute case. Witnesses who viewed same-foils lineups made slightly more low confidence false identifications in the absolute case (.10) compared to the relative case (.04). Witnesses made slightly fewer high confidence false identifications in different-foils (.07) compared to same-foils lineups (.11), possibly because the higher foil-to-suspect similarity in different-foils lineups decreased confidence in incorrect responses.

In addition to the instructions manipulation, this study also examined the effects of two foil selection methods on identification accuracy. Unlike Clark and Tunnicliff (2001) results, different-foils lineups only yielded a small nonsignificant increase in the false identification rate compared to same-foils lineups. One potential explanation for the lack of significance concerns the similarity relations between the innocent suspect, the foils, and the perpetrator. In the present study, the innocent suspect was rated as highly similar to the perpetrator ($M = 4.36$), which was substantially higher than the rating of the best-matching foil to the perpetrator ($M = 3.11$). On average, the foils used in the guilty-suspect ($M = 2.81$) and same-foils innocent-suspect lineups ($M = 2.41$) were rated as moderately similar to the perpetrator; thus, the foils may not have been adequate competitors with the innocent suspect. Also, when the perpetrator and innocent suspect are highly similar (as they were in the present study), the foils chosen for the perpetrator's lineup will also be similar to the innocent suspect. In the present study, the foils chosen for the guilty suspect's lineups were also reasonable foils for the innocent suspect in the same-foils lineups, with only slightly lower similarity on average compared to the different-foils lineup similarity relations ($M = 2.78$). In this regard, the different-foils lineups did not differ much from the same-foils lineups in terms of the similarity relations. The moderate similarity lineups coupled with the high-similarity innocent suspect may have attenuated the effect of the foil selection manipulation by making the foils poor competitors with the innocent suspect in the same-foils lineup and making the similarity relations similar between the same-foils lineups and the different-foils lineups. Further research varying the similarity of the innocent suspect to the perpetrator as well as the similarity relations of the foils is needed to establish the generality of Clark and Tunnicliff's results.

Theoretical Implications

Despite the general lack of statistical significance for the instructions manipulation, the finding of any advantage for relative judgments and not absolute judgments is striking. These results are inconsistent with the longstanding view that absolute judgments improve identification accuracy (Wells, 1984), and inconsistent with predictions based on the WITNESS model (Clark et al., 2011). The present results are consistent with comparisons of simultaneous and sequential lineups that have shown higher diagnostic accuracy for simultaneous lineups compared to sequential lineups (Gronlund, Carlson, et al., 2013; Carlson & Carlson, 2014; Dobolyi & Dodson, 2013; Mickes et al., 2012).

Wixted and Mickes (2014) have recently proposed a theory that accounts for the simultaneous lineup advantage through comparisons between lineup members. Building on classic work on perceptual learning (Gibson, 1969), they proposed that the comparison of lineup members allows witnesses to ignore non-diagnostic features that are shared across lineup members, and to focus on the distinctive, diagnostic features. The results from Experiment 1 were consistent with Wixted and Mickes' theory to the extent that relative judgments involve comparisons between lineup members, and simultaneous lineups allow for relative judgments. This analysis implies a relative judgment advantage, which was observed, but was very small. This could occur if some participants were unable to implement the decision rule according to their instructions, or if participants did not use pure versions of absolute and relative decision rules. Specifically, it may have been difficult for participants who received absolute instructions to ignore the relative differences between lineup members, and it may have been difficult for participants who received relative instructions to ignore the absolute value of the best match. As Clark et al. (2011) noted, it becomes very difficult to distinguish between absolute and relative decision rules unless participants implement pure versions of the decision rules. Consistent with this point, Fife, Perry, and Gronlund (2014) showed that even with small deviations from pure rules, for example, with the weighting parameters at $W_A = .75$ and $W_R = .25$, rather than at 1.0 and 0.0, the predictions of the model start to converge.

Practical Implications

As mentioned, sequential lineups have been broadly recommended to law enforcement as a way to improve identification accuracy by decreasing the likelihood of false identification errors. The present results suggest that the recommendation was incorrect, because the assumption underlying sequential lineups, that preventing relative judgments improves accuracy, is incorrect. In the present study, relative judgments may have improved accuracy, albeit to a small degree. Future research is needed to establish the extent to which simultaneous and sequential lineups produce relative and absolute judgments, respectively.

Limitations and Future Directions

Experiment 1 was the first empirical study to examine the effect of absolute and relative judgments on identification performance using a novel and simple paradigm with clear operational definitions. This study represents a step towards clarifying underlying eyewitness identification decision processes and an important test of a theory that has dominated the field for 30 years, the absolute-relative distinction. The finding of a relative advantage, though failing to achieve statistical significance, contradicted the well-known and widely held view that relative judgments are “dangerous" and “dysfunctional" (Wells, 1993, p. 560).

Factors other than similarity relations discussed above that may have attenuated the statistical significance of these data were (1) the subtlety of the instructions manipulation, (2) the population sampled, (3) the timing of the instructions. Regarding the subtlety of the manipulation, the instructions were lengthy and the wording of absolute and relative decision rules were necessarily similar. The impact of the instructions may have been limited to the extent that witnesses were paying attention, comprehending well, and following instructions. On a related note, witnesses were recruited using Amazon Mechanical Turk, which crowdsources nationally, producing a sample that varied substantially in demographic characteristics. In contrast to studies that use a university student subject pool, 36.6% of the participants in Experiment 1 reported that their highest level of education was a high school diploma or less, and the mean age was 30.6, ranging from 18 to 75 years old. In addition to demographic differences that might have contributed to noise in the data, participants could complete the study in an environment and on a computer platform of their choice. No attempt was made to catalog these differences, but it is likely that some witnesses completed the experiment in distracting environments or used devices that differed in screen size and resolution.

The timing of the instructions may have attenuated the effect of instructions on identification responses. Prior to receiving the instructions, witnesses made a best match determination. This was done because the WITNESS model separates the process that generates the best match and the decision rules. Witnesses in the absolute condition may have used relative information to make their best match determination. In other words, witnesses may have already committed to their identification choice using relative processing; thus weakening the effect of the absolute instruction manipulation. Future studies will provide the instructional manipulation at the outset of the lineup to avoid conflating absolute and relative judgments.

Several research questions arise from the finding of a relative judgment advantage. Are these findings replicable? Do witnesses actually use absolute and relative processes? What are the contributions of absolute and relative processes in simultaneous and sequential lineups? Do certain configurations of the foils in the lineup facilitate the use of relative judgments? To what extent do comparisons between lineup members underlie relative judgments? Experiment 2, however, continued this investigation with a more fundamental question: Do comparisons between lineup members, hypothesized to underlie relative judgments, affect eyewitness identification accuracy?

Chapter 3: Experiment 2

A witness to a crime is shown a lineup of six individuals standing in a row—one of whom is suspected by the police of having committed the crime. The witness examines the lineup, making several comparisons among lineup members. It's either lineup member number 3 or number 4. The witness continues to compare number 3 and number 4 to each other. "It's number 4," he says.

Number 4 is, indeed, the police suspect, but not all people who are suspected by the police are guilty, so the question remains: Did the witness make a correct identification of a suspect who was guilty of the crime, or did he make a false identification of a suspect who was innocent of the crime?

This hypothetical case frames the central question of Experiment 2: What effect does the comparison process have on the accuracy of eyewitness identification decisions? The results of Experiment 1 suggest that use of relative judgments may increase accuracy. Does the comparison process hypothesized to underlie relative judgments account for the performance improvement observed in Experiment 1? The present study addresses this issue with a novel paradigm designed to influence the comparison process through the positioning of lineup foils relative to the suspect.

Comparative Processes

Wixted and Mickes (2014) proposed that the comparisons between lineup members improve accuracy in simultaneous lineups by allowing witnesses to ignore shared non-diagnostic features, and focus on the unique, diagnostic features. For example, a suspect and foils may bear some resemblance to each other (e.g. same race, height, hair color). Their overlapping features are non-diagnostic of guilt. Some features, though, are not shared (e.g. face shape, eyebrow height). These features distinguish the perpetrator from the foils and the innocent suspect and thus, are diagnostic. According to Wixted and Mickes, witnesses weight the features of the lineup, then aggregate the weights and compare them to a criterion to reach a decision. In a non-comparison condition, like sequential lineups, the witness might weight diagnostic and non-diagnostic features equally. Thus, both diagnostic and nondiagnostic information contributes to the decision, which increases the match of the innocent suspect to the witness' memory of the perpetrator, increasing the false identification rate. In this case, diagnostic information may be weighted less compared to the comparison condition, decreasing the correct identification rate.

In a comparison condition, like simultaneous lineups, the witness should give more weight to the diagnostic features than the non-diagnostic features. For example, if all lineup members are white with brown hair, the witness should give little or no weight to race and hair color. Diagnostic information primarily contributes to the decision in the comparison condition—and weighting the features shared only by the guilty suspect and himself increases the correct identification rate. The focus on diagnostic features decreases the false identification rate compared to the non-comparison condition, as the innocent suspect differs from the guilty suspect in appearance. In the aggregate, decreasing the false identification rate and increasing the correct identification rate in the comparison condition results in higher accuracy compared to the non-comparison condition.

As mentioned, the theory was developed and applied to the simultaneous-sequential debate. The impact of comparisons among lineup members on identification performance, however, is broadly relevant to decision-making in eyewitness identification. It is important to examine the effect of comparisons among lineup members through many different lenses. Thus, Experiment 2 used a different experimental paradigm, inspired by studies by Tulving (1981) and by Gonzalez, Davis, and Ellsworth (1995) to investigate the role of comparison processes in eyewitness identification.

Tulving (1981) presented participants with a series of photographs, denoted A, B, C, and so on, followed by two-alternative forced choice recognition test trials that either consisted of a studied target photograph and a distractor that was very similar to that target photograph, denoted A/A', or a study photograph and a distractor that was very similar to another target photograph that was not included in that test trial, denoted A/B'. Recognition memory was more accurate for A/A' trials than for A/B' trials. One explanation of this result, suggested by Tulving, was that the comparison of A/A' allowed participants to disregard the features that were common to the target and distractor and focus on those that distinguished between the target and distractor. This account of the results, which is consistent with the model proposed by Wixted and Mickes (2014), suggested that comparisons between lineup members should increase the diagnostic accuracy of eyewitness identification evidence.

In contrast to the Tulving study, a study by Gonzalez et al. (1995) provided evidence that comparisons among lineup members can be a source of bias. They provided participants with a verbal description of a target person and a six-person lineup that included that target person. The participant's task was to indicate which of the lineup members was the best match to the description. The target person was flanked either by two lineup members who were very similar to the target or by two lineup members who were very dissimilar to the target. It is important to note that the foils were the same in the two conditions; only their positions varied. The results showed that the target person was chosen as the best match more often when surrounded by low-similarity foils than when surrounded by high-similarity foils. They explained the result as a response bias created by the contrast of low-similarity foils. Such a bias is consistent with the basic idea that comparisons between lineup members lead witnesses to identify the person who looks most like their memory of the perpetrator relative to the other members of the lineup.

However, there are a number of limitations of the Gonzalez et al. (1995) study for the present purpose. The task was not a memory task, there was no comparison of guilty versus innocent lineup conditions, and participants did not have a none-of-the-above response option. These limitations were addressed in Experiment 2, described below.

Participants became witnesses to a staged crime, followed by either a guilty-suspect lineup or an innocent-suspect lineup. Each lineup included a suspect plus two higher-similarity and two lower-similarity foils. The critical manipulation was whether the suspect was flanked by the two higher-similarity foils or by the two lower-similarity foils.

Predictions

Presumably, comparisons among lineup members will be easier when they are side-by-side than when they are separated, with other lineup members in between. If comparisons between adjacent lineup members allow witnesses to focus on diagnostic features then accuracy should increase for the higher-similarity neighbor condition over the lower-similarity neighbor condition. On the other hand, if such comparisons produce a contrast effect that biases witnesses toward an identification of the suspect, then the overall suspect identification rate may increase, with no increase in accuracy.

There is another aspect of the Tulving (1981) study that is relevant to the present study. Although the A/A' condition showed higher accuracy than the A/B' condition, it also showed lower confidence than the A/B' condition. Although confidence and accuracy are strongly correlated (Sporer, Penrod, Read, & Cutler, 1995) and calibrated (Juslin et al., 1996; Palmer, Brewer, Weber, & Nagesh, 2013) results from Tulving and others (Ross, Benton, McDonnell, Metzger, & Silver, 2007) suggest that variations in lineup composition may weaken that relationship. Specifically, comparisons among high-similarity foils may increase accuracy, but reduce confidence.

Although much research has examined the composition of the lineup (see Clark, Rush, and Moreland (2013b) for a review), there has been much less research that has examined the position of the lineup members, and none, aside from the Gonzalez et al. (1995) study, that has examined the positioning of the foils relative to the suspect. This is important because accuracy and bias may depend not only on the individuals in the lineup, but also the arrangement of those individuals.

Method

Participants

Nine hundred and ninety one undergraduate students from the University of California, Riverside participated to obtain research credit for an introductory psychology course (891, 89.9%) or a remuneration of $8.00 (103, 10.5%). (The number of participants paid and credited sums to 994 because three people were inadvertently paid and credited.) Participants were 19.3 (SD=2.1) years old on average, and 60.0% were female. The demographic breakdown of participants was: 44.2% Asian/Pacific Islander, 34.6% Hispanic, 13.9% Caucasian, 4.8% African American, and 2.4% specified that they self-identified as multiple ethnicities. Twenty-six (2.6%) of those classified as “Caucasian" self-identified as Middle Eastern.

Materials and Procedure

Participants, in groups of one to four, watched a video depicting a robbery at an ATM. The same staged crime video in Experiment 1 was used in Experiment 2. After the video, participants provided information about the hair color, height, weight, and age of the perpetrator in response to specific questions (e.g. Please give us your best estimate of his exact age). They also wrote an open ended description of the perpetrator’s appearance (e.g. Please write down everything else you remember about the description of the robber). Participants rated their confidence that they would be able to identify the perpetrator if they saw him again on a 1-to-6 scale (prospective confidence). Participants then completed a series of filler tasks, including a school aspirations questionnaire, a short version of the Big Five Personality Inventory (John, Donahue, & Kentle, 1994) and a series of multiplication problems. These three tasks created a 15-minute delay between the crime video and the presentation of a five-person photo lineup. The composition of the lineup defined the primary experimental manipulation.

Table 7
Design and composition of high- and low-similarity neighbor lineups by suspect position. S denotes the suspect, H_1 and H_2 denote high-similarity foils, and L_1 and L_2 denote low-similarity foils.

	Suspect Position		
	Position 2	Position 3	Position 4
High-Similarity	H_1 S H_2 L_1 L_2 H_2 S H_1 L_2 L_1	L_1 H_1 S H_2 L_2 L_2 H_2 S H_1 L_1	L_1 L_2 H_1 S H_2 L_2 L_1 H_2 S H_1
Low-Similarity	L_1 S L_2 H_1 H_2 L_2 S L_1 H_2 H_1	H_1 L_1 S L_2 H_2 H_2 L_2 S L_1 H_1	H_1 H_2 L_1 S L_2 H_2 H_1 L_2 S L_1

Lineup Materials and Construction. The design and composition of the lineups are illustrated in Table 7. The lineups appear in Figure 7. Each lineup consisted of a guilty or innocent suspect plus two higher-similarity foils H_1 and H_2 and two lower-similarity foils L_1 and L_2. The same guilty suspect used in Experiment 1 was used in Experiment 2. A different innocent suspect who was rated as having only moderate similarity to the guilty suspect was used in Experiment 2. In the high-similarity neighbor condition the suspect was flanked by the two higher-similarity foils, whereas in the low-similarity neighbor condition the suspect was flanked by the two lower-similarity neighbors. The lineups were presented in a single horizontal line to simplify the relationship between the suspect and the neighboring foils. In a single-row, five-person lineup, the suspect can have only two neighbors, one to the left and one to the right. The suspect was presented equally often in positions 2, 3, and 4. (The neighborhood manipulation required that the suspect not appear in positions 1 or 5.) The positions of the foils relative to the suspect were also counterbalanced so that the positions of the two higher-similarity foils and the two lower-similarity foils appeared

Figure 7. *Lineups from Experiment 2. Top two lineups are guilty suspect lineups, bottom two lineups are innocent-suspect lineups. First and third lineups represent the high-similarity condition, second and fourth lineups represent the low-similarity condition. The suspect appears in the third position in each.*

equally often on either side of the suspect (H_1 to the left and H_2 to the right versus H_1 to the right and H_2 to the left; and likewise for L_1 and L_2). The lineups were created to equate foil-to-suspect similarity for guilty-suspect and innocent-suspect lineups.

Mean ratings of the appearance of the foils in the innocent suspect lineup to the appearance of the guilty suspect are shown in Table 8, in the sub-column labeled *To Sg* (to guilty suspect). An assumption of the WITNESS model is that the foils in the innocent suspect lineup are not more similar to the perpetrator than to the innocent suspect. Because these foils were chosen based on their similarity to the innocent suspect, the innocent suspect's foils should share more features with him than with the guilty suspect. In this study, this assumption was met for three of the four foils, but good-foil 1 was slightly more similar to the guilty suspect (mean = 2.20) than the innocent suspect was to the guilty suspect (mean = 2.02).

Table 8

Mean similarities of lineup members to guilty and innocent suspects. I_s stands for innocent suspect and G_s stands for guilty suspect. The ratings in the To I_s column are ratings of the similarity of a given lineup member to the innocent suspect, and the ratings in the To G_s column are ratings of the similarity of a given lineup member to the guilty suspect.

	Guilty-Suspect	Innocent-Suspect	
Member		To I_s	To G_s
Suspect	–	3.13	2.02
HS-Foil 1	3.47	3.56	2.20
HS-Foil 2	3.22	3.16	2.00
LS-Foil 1	1.27	1.27	1.10
LS-Foil 2	1.37	1.40	1.19

Lineup procedure. The instructions (presented in writing and by the experimenter) directed participants to look at all the photographs and indicate whether the person who committed the crime was in the lineup. Participants were instructed that the person who committed the crime may or may not appear in the lineup. Participants recorded their decision by circling the number corresponding to one of the photographs or by checking a box to indicate perpetrator absence. They indicated their confidence in their decision on a 1-6 scale.

After the lineup all participants rated the similarity of 10 new mugshot photographs to their memory of the person in the video on a scale of 1 ("Looks very little like your memory of the person in the video") to 6 ("Looks very much like your memory of the person in the video"). The mugshots were of individuals who ranged from high to low similarity to the perpetrator's appearance and were in the same ethnic group as the perpetrator. After the similarity ratings, participants filled out a demographic questionnaire and answered questions about the strategy they used to make their decision and viewing conditions. Participants were given a consent and debriefing form that described the purpose of the study and thanked for their participation.

Results

Nine hundred and twelve participants were included in the following three sets of analyses[1]. The first set compared identification performance in high- and low-similarity neighbor conditions for both guilty and innocent suspect lineups. The second set of analyses examined differences in discriminability by neighborhood condition via d' and ROC analyses. This set of analyses directly addressed the question: How does the comparison process affect eyewitness accuracy? The third set of analyses investigated the relationship between confidence and accuracy, and was also relevant to the interpretation of the d' and ROC analyses.

Identification Responses

Table 9 shows the response proportions for suspect, foil, and non-identification responses for high-similarity neighbor and low-similarity neighbor conditions, as a function of suspect guilt and suspect position. These data were statistically evaluated through separate log-linear analyses for guilty and innocent-suspect lineups, for each response type. As in Experiment 1, likelihood ratio chi-square statistics are reported for the loglinear analyses, and Pearson's chi-square statistics are reported for focused follow-up analyses.

Guilty Suspect lineups.

Correct identifications. The correct identification rate was higher when the guilty suspect was surrounded by higher-similarity foils (.59) than when surrounded by lower-similarity foils (.50), $\chi^2(1, N = 456) = 3.907$, $p = .048$, $r = .093$. The correct identification rate varied slightly, but not significantly, as a function of the suspect position, $\chi^2(2, N = 456) = 4.217$, $p = .121$. The highest-order interaction in the log-linear model (identification response × neighborhood × suspect position) was not statistically significant, indicating that the neighborhood effect did not vary with the position of the suspect, $\chi^2(2, N = 456) = 1.052$, $p = .591$.

---

[1] Of the 991 original participants, 79 (8%) were excluded from the analyses, for reasons as follows: five (.05%) were inadvertently exposed to the staged crime prior to their participation, two (.02%) experienced a computer malfunction, three (.03%) participated twice (data from their second time was dropped), 51 (5%) participants provided unclear responses on the identification task (by making an identification and rejecting the lineup), one (.01%) participant requested help multiple times from the experimenter and reported that he could not answer the identification questions, 10 (1%) because of experimenter error, and seven (.07%) for omitting responses on the identification response form.

Table 9
Response frequencies and proportions (in parenthesis) for responses broken up by lineup type, neighborhood, and suspect position.

		Guilty Suspect			Innocent Suspect		
		High Similarity	Low Similarity	p	High Similarity	Low Similarity	p
Suspect Position 2	Suspect	51 (.671)	42 (.553)	.13	26 (.342)	22 (.289)	.49
	Foil	17 (.224)	15 (.197)	.69	19 (.250)	29 (.382)	.08.
	No Pick	8 (.105)	19 (.250)	.02*	31 (.408)	25 (.329)	.31
Suspect Position 3	Suspect	43 (.566)	33 (.434)	.10.	17 (.224)	17 (.224)	1.0
	Foil	20 (.263)	24 (.316)	.47	17 (.224)	22 (.289)	.35
	No Pick	13 (.171)	19 (.250)	.23	42 (.553)	37 (.487)	.42
Suspect Position 4	Suspect	41 (.539)	39 (.513)	.75	23 (.303)	27 (.355)	.49
	Foil	21 (.276)	21 (.276)	1.0	18 (.237)	22 (.289)	.46
	No Pick	14 (.184)	16 (.211)	.68	35 (.461)	27 (.355)	.19

Foil identifications. The foil identification rate was slightly, but not significantly, lower for the high-similarity neighbor condition (.58) than the low-similarity neighbor (.60) condition, $\chi^2(1, N = 456) = 0.046$, $p = .830$, $r = .010$. The foil identification rate did not vary reliably across suspect positions, $\chi^2(2, N = 456) = 2.900$, $p = .235$. No other effects were statistically significant ($.830 < p < 1.00$).

Lineup rejections. The incorrect rejection rate was significantly lower in the high-similarity neighbor condition (.15) than in the low-similarity neighbor condition (.24), $\chi^2(1, N = 456) = 5.072$, $p = .024$, $r = .105$. Neither the response × suspect position interaction nor the highest-order interaction between neighborhood, response, and suspect position were statistically significant, $\chi^2(2, N = 456) = 0.533$, $p = .766$ and $\chi^2(2, N = 456) = 2.104$, $p = .349$, respectively. The latter finding indicates that the effect of neighborhood on lineup rejection rates did not vary with suspect position.

Innocent suspect lineups.

False identifications. The false identification rates were identical (.289) in high-similarity neighbor and low-similarity neighbor lineups, $\chi^2(1, N = 456) = 0.0$. The response × suspect position interaction was marginally significant, $\chi^2(2, N = 456) = 4.992$, $p = .082$. Follow-up pairwise analyses showed that the suspect identification rate was significantly higher when the suspect was in position 4 (.33) compared to position 3, (.22), $\chi^2(1, N = 304) = 4.211$, $p = .040$, $r = .118$, marginally higher when the suspect was in position 2 (.31) compared to 3 (.22), $\chi^2(1, N = 304) = 3.273$, $p = .070$, $r = .104$, and nonsigificantly higher when the suspect was in position 4 compared to 2, $\chi^2(1, N = 304) = .060$, $p = .806$, $r = .014$. In other words, the innocent suspect was less likely to be identified when placed in the middle of the lineup (position 3) than when placed slightly to the left (position 2) or slightly to the right (position 4). The response × neighborhood × suspect position interaction was not statistically significant, $\chi^2(2, N = 456) = 0.965$, $p = .617$, indicating no interaction between neighborhood similarity and suspect position.

Foil identifications. The foil identification rate was lower in the high-similarity neighbor condition (.24) than in the low-similarity neighbor condition (.32), $\chi^2(1, N = 456) = 3.951$, $p = .047$, $r = .093$. No other effects were statistically significant ($.455 > p > .993$).

Lineup rejections. The loglinear analysis showed that correct rejections of the lineup were slightly but not significantly higher when the innocent suspect was surrounded by high-similarity neighbors (.47) compared to low-similarity neighbors (.39), $\chi^2(1, N = 456) = 3.230$, $p = .072$, $r = .084$. The interaction between lineup rejection responses and suspect position was statistically significant, $\chi^2(2, N = 456) = 7.620$, $p = .022$. Follow-up pairwise comparisons showed that the innocent-suspect lineup was correctly rejected significantly more often when the suspect appeared in position 3 (.52) compared to 2 (.38) and marginally more often when the suspect appeared in position 3 compared to 4 (.41), $\chi^2(1, N = 304) = 7.049$, $p = .008$, $r = .152$, and $\chi^2(1, N = 304) = 3.823$, $p = .051$, $r = .112$, respectively. Participants correctly rejected the lineup slightly, but not significantly, more often when the suspect appeared in position 4 compared to 2, $\chi^2(1, N = 304) = 0.499$, $p = .480$, $r = .041$. In other words, the innocent suspect lineup was rejected more often when the suspect appeared in the middle of the lineup (position 3) compared to the outside positions. No other effects were statistically significant ($.932 < p < .973$).

Discriminability

The next set of analyses evaluated the ability of the witnesses to discriminate between suspects who were guilty versus suspects who were innocent as a function of neighborhood similarity and suspect position using d' and pAUC.

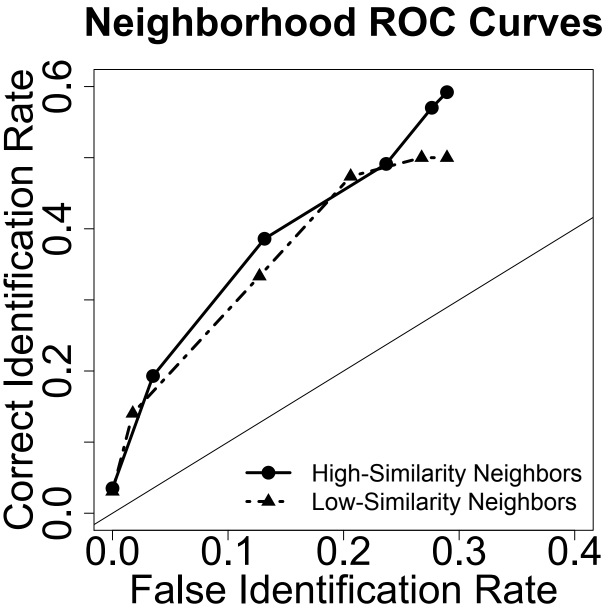


Figure 8. Receiver Operating Characteristic (ROC) plot for high- and low-similarity neighbor conditions. The diagonal line represents chance performance.

Accuracy was higher for the high-similarity neighborhood lineups ($d' = 0.788$) compared to the low-similarity neighborhood lineups ($d' = 0.555$), although the difference was not statistically significant, $G = 1.360$, $p = .174$. The results of the ROC analysis are shown in Figure 8 for high- and low-similarity neighbor conditions. The pAUC was larger for the high-similarity neighborhood condition (.172) than for the low-similarity neighborhood condition (.157), but the difference was not statistically significant, $D = .810$, $p = .418$.

The d' and ROC analyses both showed a similar pattern: a small, but statistically non-significant discriminability advantage when the suspect was surrounded with high-similarity foils rather than low-similarity foils. The close correspondence between d' and pAUC is expected to the extent that they arise from a common signal detection theory framework (Mickes, et al., 2014).

The advantage for high-similarity neighbors, although small, was consistent across suspect positions. In such cases, where effects are small but consistent, it can be illuminating to meta-analytically aggregate results across conditions. The d' and pAUC values were computed for each suspect position. A one-sample t-test compared the high-similarity and low-similarity neighbor conditions. The d' and pAUC values are shown in Table 10. For these analyses, the d' differences were statistically significant, $t(2) = 4.718$, $p = .042$, $r = .958$, but the pAUC differences were not, $t(2) = 2.077$, $p = .173$, $r = .827$. This small discrepancy between d' and pAUC will be discussed later in this chapter.

Table 10
d' and pAUC, across suspect positions for high-similarity and low-similarity neighbors conditions.

	Suspect Position			
	Position 2	Position 3	Position 4	Overall
d' High-Similarity	0.850	0.925	0.616	0.788
d' Low-Similarity	0.687	0.594	0.404	0.555
pAUC High-Similarity	0.180	0.182	0.158	0.172
pAUC Low-Similarity	0.158	0.155	0.157	0.157

The identification results can be summarized as follows: Identification accuracy improved in several ways when high-similarity neighbors surrounded the suspect. Witnesses were more likely to correctly identify the guilty suspect and no more likely to identify the innocent suspect when the suspect was surrounded by high-similarity neighbors compared to low-similarity neighbors. Fewer witnesses identified foils when high-similarity neighbors surrounded the innocent suspect (and to a very small degree, the guilty suspect) compared to when low-similarity neighbors surrounded the innocent suspect. Further, more witnesses correctly rejected the innocent suspect lineup when the innocent suspect was surrounded by high-similarity compared to low-similarity neighbors. The statistically significant advantage for the correct identification rate and equal false identification rates in the high- versus low-similarity neighbor conditions translated to a small discriminability advantage indicated by a statistically significant increase in d' for high-similarity neighbors. The pAUC analyses showed the same pattern, but the advantage for high-similarity neighbors did not reach statistical significance. What accounts for the statistical differences between the correct and false identification rates, the d' analyses, and the pAUC analyses? More to the point, how can the correct identification rate be significantly higher and the false identification rates be identical–and yet the differences in pAUC not be statistically significant? The next section addresses this question by examining the relationship between confidence and accuracy for high- and low-similarity neighbor conditions.

Confidence and the confidence-accuracy relationship

Confidence data are useful in understanding underlying mechanisms, and in the present case, the relationship between confidence and accuracy may provide an explanation for the discrepancies between pAUC and d' measures of accuracy. Assuming that the increase in the correct identification rate in the high-similarity neighbor condition was produced by an increase in the match to memory for the guilty suspect, then there should also be an increase in confidence. Confidence should be higher in the high-similarity neighbor condition than in the low-similarity neighbor condition. However, average confidence for correct identifications was slightly (but not significantly) lower for the high-similarity neighbor condition (3.83) than for the low-similarity neighbor condition (3.96), $t(247) = .906$, $p = .366$, $r = -.058$. Average confidence for false identifications was slightly (but not significantly) higher for the high-similarity neighbor condition (3.35) than for the low-similarity neighbor condition (3.14), $t(130) = 1.160$, $p = .248$, $r = .101$, which seems consistent with results showing identical false identification rates.

Of course, the confidence means can be the same even if the distributions of confidence are different. The distributions of confidence for suspect identifications from guilty and innocent suspect lineups are shown in Figure 9. The distributions of confidence for the high-similarity neighbor and low-similarity neighbor lineup conditions were significantly different for guilty-suspect lineups, $\chi^2(5, N = 249) = 12.511$, $p = .028$, but not for innocent-suspect lineups, $\chi^2(4, N = 132) = 3.969$, $p = .410$. Focusing on the guilty-suspect lineups, the confidence distribution for the high-similarity neighbor condition showed more low-confidence suspect identifications, more high-confidence suspect identifications, and fewer moderate confidence suspect identifications, relative to the low-similarity neighbor condition.

The relationship between confidence and accuracy was assessed by calculating point-biserial correlations and by plotting the calibration curves. The point-biserial correlations, considering only positive identifications, were $r = .243$ ($N = 313$) for high-similarity and $r = .316$ ($N = 313$) for

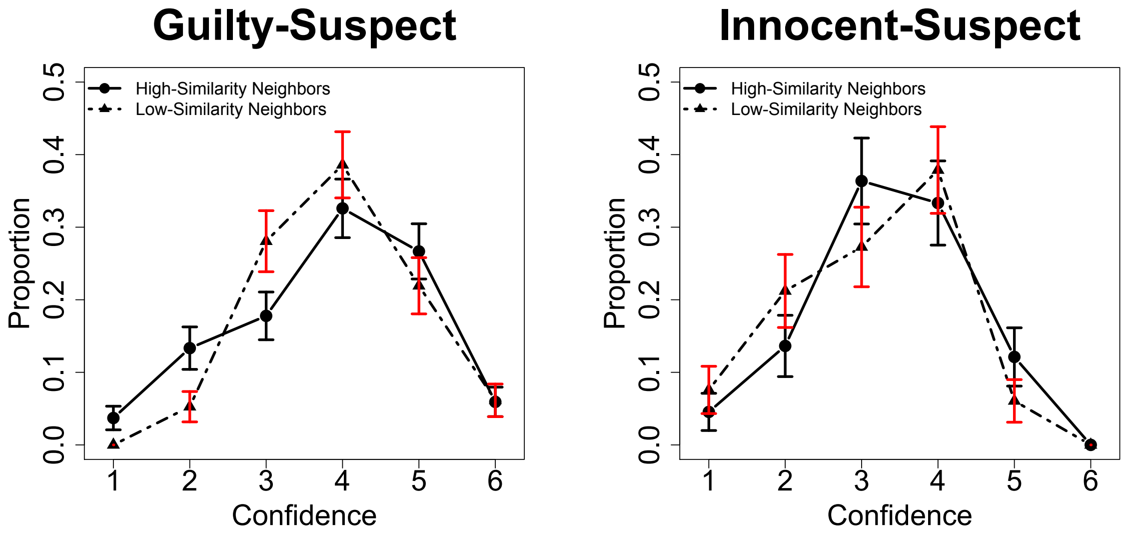


Figure 9. *The distributions of confidence for the high-similarity neighbor and low-similarity neighbor lineup conditions for Guilty-Suspect Lineups (left) and Innocent-Suspect Lineups (right). Error bars represent standard error of a proportion. Error bars for high-similarity neighbors appear in black, error bars for low-similarity neighbors appear in red.*

low-similarity neighbor conditions. Both of these correlations were significantly greater than zero, $p < .001$, and not significantly different from each other, $z = .986$, $p = .322$. Considering only suspect identifications, the correlations were .191 for high-similarity neighbors ($p = .007$) and .365 for low-similarity neighbors ($p < .001$). The difference between the two correlations was marginally significant ($p = .067$).

The confidence-accuracy calibration curves are shown in Figure 10. It is clear from the Figure that confidence and accuracy were highly calibrated for low-similarity neighbor lineups, and relatively less well calibrated for high-similarity neighbor lineups. The correlation between confidence and accuracy was $r = .768$, for high-similarity neighbors and $r = .970$ for low-similarity neighbors. Even with only six data points in each calibration curve, these high correlations were marginally significant ($p = .074$ with high-similarity neighbors), or highly significant ($p < .001$ with low-similarity neighbors). Although the two correlations are not statistically different from each other ($z = 1.319$, $p = .187$), it is clear from Figure 10 that the calibration curve for the low-similarity neighbor condition is linear whereas the calibration curve for the high-similarity neighbor condition is slightly concave.

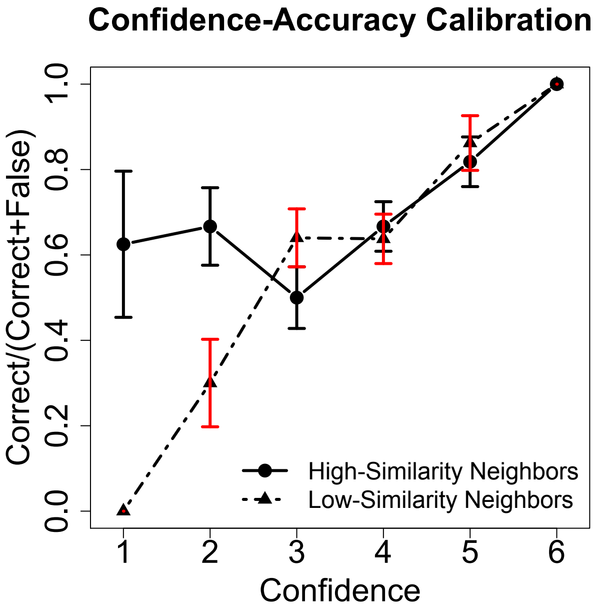


Figure 10. Confidence-accuracy calibration for high- and low-similarity neighbor conditions. Error bars represent standard error of a proportion. Error bars for high-similarity neighbors appear in black, error bars for low-similarity neighbors appear in red.

The concavity of the high-similarity neighbor calibration curve is due to a higher proportion of low-confidence correct identifications compared to the low-similarity neighbor condition. Seventeen percent of participants who correctly identified the guilty suspect when he was surrounded by high-similarity neighbors did so with very low confidence (judgments of “1" or “2"). The proportions of suspect identifications that were correct for those who rated their confidence as a “1" or a “2" were .625 and .667 in the high-similarity neighbors condition and .000 and .300 in the low-similarity neighbors condition, respectively. Also, for the high-similarity neighbors condition, the accuracy associated with confidence ratings of “1" and “2" was higher than that associated with confidence ratings of “3" (.500).

It should be noted that the concavity of the high-similarity neighbor calibration curve could be due to a low number of observations. There were 5 correct and 3 false identifications made in the high-similarity neighbor condition and 0 correct and 5 false identifications made in the low-similarity neighbor condition at a confidence of 1. A chi-square test comparing correct and false identifications made with confidence of 1 for high and low-similarity neighbors was statistically significant, $\chi^2(1, N = 13) = 4.800$, $p = .029$, $r = .608$, although the low number of observations are problematic for this analysis. There were 18 correct and 9 false identifications made in the high-similarity neighbor condition and 6 correct and 14 false identifications made in the low-similarity neighbor condition at a confidence of 2. A chi-square test comparing correct and false identification responses for high and low-similarity neighbors combining responses made with confidence of 1 and 2 was statistically significant, $\chi^2(1, N = 60) = 10.740$, $p = .001$, $r = .423$, indicating that there were more correct identifications made at lower levels of confidence for high-similarity neighbors compared to low-similarity neighbors.

The shape of the ROC curves also reflected the additional low-confidence correct identifications in the high-similarity neighbor condition. Figure 8 shows a very small advantage for the high-similarity neighbor condition at the high-confidence end of the ROC curve, which disappears for moderate levels of confidence, and then reappears for the low confidence end of the ROC. Note, in particular, that the rightmost points of the ROC curve for the high-similarity condition bend sharply upward whereas the rightmost points of the ROC curve for the low-similarity condition flatten.

The differences in the confidence-accuracy relationships for high-similarity and low-similarity conditions provide a tentative explanation for why the high-similarity advantage was statistically non-significant in the pAUC analysis, even though the correct identification rate was significantly higher and the false identification rate dead equal to that in the low-similarity condition. The area under the curve increases to the extent that correct identifications are made with higher confidence than false identifications. Thus, it could be the case that if the low-confidence correct identifications in the high-similarity neighbor condition had been made with higher confidence, the pAUC advantage would have been statistically reliable.

One way to evaluate this speculation is to create a distribution of confidence judgments for correct identifications in the high-similarity neighbor condition that is shifted upward relative to the low-similarity condition, implementing a core assumption of signal detection theory and ROC analysis. With this assumption the statistical comparison of pAUC showed a significant advantage for the high-similarity neighbor condition (pAUC = .194) over the low-similarity neighbor condition (pAUC = .157), $D = 2.024$, $p = 0.043$. It is important to be very clear about the purpose of this exercise. The purpose was not to replace actual data with hypothetical data in order to suggest that the pAUC analysis really did show a statistically significant difference. Clearly, it did not. Rather, the purpose was to demonstrate how statistical conclusions based on pAUC can vary depending on the relationship between confidence and accuracy.

Discussion

Experiment 2 examined the effects of making comparisons between lineup members on the accuracy of eyewitness identification decisions. The high-similarity neighbor condition was designed to facilitate comparisons among the highest-similarity lineup members by surrounding the suspect with high-similarity foils, allowing witnesses to more easily compare the three best-matching lineup members to identify and compare diagnostic features. In contrast, the low-similarity neighbor condition surrounded the suspect with low-similarity foils, making it more difficult to make comparisons among the best matching lineup members by separating them.

Surrounding the suspect with high-similarity neighbors resulted in a higher correct identification rate and a lower false rejection rate in guilty suspect lineups, and a higher correct rejection rate and a lower foil identification rate in innocent suspect lineups. False identification rates of the innocent suspect were exactly equal in the high- and low-similarity neighbor conditions. In this regard, a manipulation designed to facilitate the witness's ability to focus on diagnostic features and discount shared features yielded more accurate identification decisions. Phrased slightly differently, these results suggest that comparisons among high-similarity alternatives increased accuracy, whereas comparisons among low-similarity alternatives decreased accuracy.

The d' and ROC analyses both showed a small advantage in discrimination between guilty and innocent suspects for high-similarity neighbors. The advantage was not statistically significant in the analyses that aggregated across suspect positions, precluding an open-and-shut conclusion that comparative processes improve accuracy. However, when accuracy measures were computed by suspect position and combined meta-analytically, d' analyses showed a statistically significant accuracy advantage for high-similarity neighbors, although a similar analysis for pAUC did not. It is possible that the discrepancy between d' and pAUC analyses was due to the relationship between confidence and accuracy. A follow-up analysis that shifted the confidence distribution upward for the high-similarity neighbors condition, lent support to this reasoning.

The present results are consistent with the confidence-accuracy inversion obtained in Tulving's (1981) study, which showed both higher accuracy and lower confidence when the target was tested with a similar distractor (A/A') relative to when the target was tested with a distractor that was similar to some other studied item (A/B'). In the present study, accuracy was higher for the high-similarity condition, and confidence was slightly although not significantly lower relative to the low-similarity condition. This affected the shape of the ROC curves and the statistical conclusions based on the ROC analyses, although the direction of the effect remained the same (the high-similarity neighbors ROC was always higher than the low-similarity neighbors ROC). A reanalysis that shifted the confidence distribution for the high-similarity neighbor condition upward, consistent with signal detection theory, did produce a significant accuracy advantage for the high-similarity neighbor condition. It is important to note that the reanalysis did not change any identification responses, but only the confidence of correct identifications in the high-similarity condition. The purpose of this reanalysis was not to fudge the data in order to get a statistically significant outcome, but rather to assess the extent to which a disconnection in the relationship between confidence and accuracy might affect estimates of pAUC and the statistical conclusions based on those estimates.

Theoretical Implications

The results of the present study showed a small overall increase in accuracy for the high-similarity neighbors condition, consistent with Wixted and Mickes' (2014) proposal that comparisons among lineup members increase accuracy by allowing witnesses to focus on diagnostic features and ignore non-diagnostic features of lineup members, and inconsistent with the long-standing view that comparisons between lineup members increase the likelihood of error and decrease the overall accuracy of eyewitness identification evidence (Wells, 1984; Lindsay & Wells, 1985; Steblay et al., 2001; Wells, Steblay, & Dysart, 2014; Dunning & Stern, 1994). The results also provide no support for the bias model, which predicts that guilty and innocent suspect identifications should both increase when the suspect is surrounded by low-similarity neighbors. Correct identifications did; false identifications did not.

As in Experiment 1, the results of Experiment 2 are consistent with research comparing simultaneous and sequential lineups, showups and lineups, and lineups with more or less-similar foils (Carlson & Carlson, 2014; Dodson & Dobolyi, 2013; Gronlund et al. 2013; Mickes et al. 2012). Comparisons between lineup members are impossible in showups because there are no foils; thus, as with the recent simultaneous-sequential findings, the condition that prohibits comparisons between lineup members is associated with lower eyewitness identification accuracy. The current study provides a complementary conclusion—that comparative processes are indeed operative in simultaneous lineups and may be facilitated depending on the lineup composition. In keeping with the latter observation, the present results are also consistent with meta-analyses showing that lineups composed of higher similarity foils improve accuracy relative to those composed of low-similarity foils (Clark, 2012; Fitzgerald, Price, Oriet, & Charman, 2013).

The results also have important theoretical implications regarding the relationship between confidence and accuracy. Signal detection models assume that recognition decisions and confidence judgments are based on the same underlying variable; that variable may be described in terms of the strength of the memory trace, or the match between test items and memory, but whatever term is used to describe it, it is the basis for both recognition decisions and confidence. Confidence is thought to indicate the location of the recognition response on the memory strength (decision) axis (Wickens, 2001), with higher confidence responses indicating higher memory strength. In this manner, confidence judgments provide a finer partitioning of the strength of the participant's memory for a given test stimulus.

Consistent with this assumption, the relationship between confidence and accuracy is typically very strong, not only for lists of verbal materials (Murdock, 1974) but also for eyewitness identification lineups (Sporer et al., 1995; Wixted et al., In Press). Although confidence and accuracy were on the whole strongly correlated and calibrated in the present study, their relationship was complicated by the fact that the high-similarity neighbor condition showed higher accuracy, but not higher confidence, relative to the low-similarity neighbor condition, and also showed a slightly concave calibration curve.

If the comparison process increases accuracy by focusing witnesses on diagnostic features, why does this not increase their confidence in their identifications? One possibility is that participants may make assumptions about their likelihood of being accurate based on a subjective feeling of difficulty (Busey, Tunnicliff, Loftus, & Loftus, 2000; Van Zandt, 2000). Close comparisons between similar choice alternatives may increase accuracy, but make the task seem more difficult to the extent that one has to focus more carefully in order to determine which features are diagnostic. Alternatively, the comparison process may increase accuracy not by increasing the match of the guilty suspect to the witness's memory of the perpetrator, but rather by decreasing the match values of the otherwise closely matching foils. To the extent that witnesses base their confidence on the value of the best match, this would increase accuracy, by allowing witnesses to rule out the foils, but have no effect on confidence.

Methodological Implications

ROC curves are useful for assessing the diagnostic accuracy of eyewitness identification procedures (Mickes, et al., 2012, Wixted & Mickes, 2012). An important assumption for ROC curves constructed from confidence judgments is that confidence judgments reflect the underlying memory strength or match of the test stimulus. That assumption may have been violated in the present study, raising a question as to the robustness of such analyses to such violations. This does not mean that confidence judgments cannot be used for ROC analyses of eyewitness identification data. The disconnection between confidence, accuracy, and memory strength in the present study did not alter the general pattern—the partial area under the curve was higher for the high-similarity neighbor condition than for the low-similarity neighbor condition, consistent with the d' analyses and consistent with results showing a higher correct identification rate and equivalent false identification rates. However, the analyses showed that the confidence-accuracy relationship can, and did, have an effect on the statistical conclusions. This suggests that ROC analyses based on confidence judgments should be considered carefully, and interpreted in the light of the confidence distributions and the confidence-accuracy relationship.

Practical Implications

Although many studies have shown that suspect position can affect identification decisions (Clark & Davey, 2005; Gronlund et al., 2009; Mecklenburg, 2006; O'Connell & Synnott, 2009) only one study, by Gonzalez et al. (1995), examined the positioning of the foils relative to the suspect, and as noted earlier, that study looked at measures of lineup bias, not eyewitness identification accuracy. It is important to understand how the neighborhood of the suspect might affect identification confidence and accuracy in actual criminal investigations.

The present results are at odds with policy recommendations favoring sequential over simultaneous lineups. The sequential lineup procedure was developed to prevent comparisons between lineup members. The results of the current study suggest that lineup procedures that inhibit comparative processes may not reduce false identifications, and may reduce correct identifications, leading to an overall decrease in the accuracy of eyewitness identification evidence.

The results also have implications for how confidence judgments are interpreted and utilized by judges and juries. The U.S. Supreme Court in *Neil v. Biggers (1972)* and *Manson v. Brathwaite (1977)* stipulated witness confidence as an index of reliability that trial courts could consider in deciding whether eyewitness evidence should be admitted at trial, and *U.S. v. Telfaire (1972)* listed witness confidence as a factor that jurors could consider in determining the weight to attach to identification evidence. These rulings are consistent with the assumption that confidence is a useful predictor of accuracy, which indeed it is. However, the results suggest that by increasing the similarity of the foils next to the suspect, witnesses may make reliable identifications, but with very low confidence. Such reliable, but low-confidence identifications may be less likely to be admitted into evidence, and may be viewed with skepticism by jurors.

Limitations and Future Research

The present study represents an important first step in understanding how the arrangement of lineup foils may influence the decision processes that underlie eyewitness identification, and how changes in those decision processes may affect both accuracy and confidence. It is, however, only a first step.

The generality of the effect has yet to be determined. How will the pattern of results be affected by variations in stimulus materials, instructions, and the conditions of observation and memory? To what extent might neighborhood effects vary across more complex configurations? In the present study, the lineup was presented in a single line in order to simplify the definition of the suspect's “neighborhood". However, in the U.S. photographic lineups are often presented in 2×3 arrays (two rows with three photographs each). Larger arrays, 3×3 and 3×4, have also been used in some jurisdictions. In such larger arrays, neighborhoods may not only include the foils on either side, but also the “upstairs" and “downstairs" neighbors as well as those who share only a corner.

The results of Experiment 2, combined with results showing a simultaneous lineup advantage, point toward a host of new questions about the nature of comparison processes, how comparisons between lineup members affect or interact with other decision processes, and how comparison and decision processes affect accuracy and confidence in eyewitness identification. Experiment 3 considered the effect of comparisons between lineup members in simultaneous and sequential lineups.

Chapter 4: Experiment 3

In recent years the California legislature has proposed several bills (such as AB 604, SB 1544, SB 1591, and SB 756) to implement “best practices" in law enforcement procedures for eyewitness identification. As recently as September 2013, the California State Senate amended AB 807 to authorize law enforcement agencies to adopt identification procedures recommended by the California Commission on the Fair Administration of Justice. According to the Commission’s review of the psychological science, these recommended procedures decrease the risk of false identification of the innocent, thus decreasing wrongful convictions. One of the recommended procedures was the sequential lineup.

The recommendation was based primarily on Lindsay and Wells’ (1985) finding that sequential lineups increased accuracy by preventing witnesses from comparing the lineup members to each other. As mentioned, current research challenged both the empirical claim that sequential lineups increased accuracy and the theoretical claim that comparisons among lineup members decreased accuracy (Mickes et al., 2012; Wixted & Mickes, 2014).

Experiment 2 showed that surrounding the suspect with high-similarity neighbors may facilitate comparisons between the best matching lineup members, thereby increasing accuracy to a small extent. Experiment 3 continued building on this foundation to address three questions of theoretical and practical importance: (1) Is the accuracy advantage for simultaneous lineups due to the comparison process? (2) Does the comparison process, and hence the simultaneous accuracy advantage, depend on the composition of the lineup, particularly the foils that surround the suspect? (3) To what extent does identification accuracy in simultaneous and sequential lineups depend on the position of the suspect in the lineup?

In Experiment 3, participants witnessed a staged crime, followed by a guilty or an innocent suspect lineup. The guilty and innocent suspect lineups were presented simultaneously or sequentially. For all lineup types, the high-similarity neighbor condition surrounded the suspect with high-similarity foils and the low-similarity neighbor condition surrounded the suspect with low-similarity foils. In sequential lineups, neighbor similarity is temporal rather than spatial. For example, in the high-similarity neighbor sequential condition, one high-similarity neighbor directly preceded the presentation of the suspect, and another high-similarity neighbor directly followed the suspect. Position of the suspect was varied systematically so that the suspect appeared in the 2nd, 3rd or 4th position equally often.

Predictions

To the extent that comparisons among lineup members underlie the simultaneous advantage as theorized by Wixted and Mickes (2014), the simultaneous lineup advantage should be larger when higher-similarity foils surround the suspect. Because sequential lineups preclude the comparative processes that are thought to underlie the benefit in the high-similarity condition, high-similarity neighbors should not increase accuracy in sequential lineups, but rather may decrease accuracy depending on the position of the suspect. The basis for this prediction comes from evidence that accuracy decreased when the suspect appeared in a late lineup position and was preceded by a high-similarity foil that drew identification responses early in the lineup (Clark & Davey, 2005).

Method

Participants

A total of 1692 participants from the same participant pool as in Experiment 2 participated in Experiment 3 to obtain research credit for an introductory psychology course (1680, 99.3%) or a remuneration of $8.00 (12, 00.7%). Participants were 19.5 (SD=1.78) years old on average, 55.6% were female, 44.1% were male, and 0.3% declined to state their gender. Participants fell into the following demographic categories: 44.8% Asian/Pacific Islander, 32.0% Hispanic, 13.9% Caucasian, 4.0% African American, 0.5% Native American, 4.5% specified that they self-identified as multiple ethnicities, and 0.3% declined to state their ethnicity. Fifty-three (3.2%) of those classified as “Caucasian" self-identified as Middle Eastern.

Materials and Procedure

The neighborhood manipulation, stimulus materials and procedures were the same as those described in Experiment 2 with the following exceptions:

1. Half of the participants were randomly assigned to a simultaneous lineup condition and the other half were assigned to a sequential lineup condition. In the sequential condition, lineup members were presented one at a time. All lineup members were presented at once in the simultaneous condition as in Experiment 2.

2. All aspects of Experiment 3 were programmed in MATLAB using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Upon beginning the experiment, participants wore headphones and entered their ethnicity, age, and gender information into the experimental program. After entering their personal information, participants watched the staged-crime video. Lineup instructions were recorded in advance of the experiment and delivered over headphones after the distractor task and prior to the lineup. Rather than writing answers on a form, as in Experiment 1, participants pressed a key corresponding to the lineup member's number to make an identification and pressed the “n" key to reject the lineup in the simultaneous condition. Each lineup member was presented separately in the center of the screen in the sequential condition. Participants either pressed the “y" key to identify the lineup member or the “n" key to reject the lineup, then typed their retrospective confidence judgment, once for each lineup member.

3. Prospective and retrospective confidence judgments were obtained with a 0-100 scale, rather than a 1-6 scale. The anchors for the prospective judgment were: 0 (I am not at all confident that I would recognize the robber if I saw him again) - 100 (I am completely confident that I would recognize the robber if I saw him again). For the retrospective confidence judgment, if the witness identified the lineup member, the anchors were: 0 (I am not at all confident that I chose the perpetrator) - 100 (I am completely confident that I chose the perpetrator). If the witness rejected the lineup the anchors were: 0 (I am not at all confident that the perpetrator was not in the lineup) - 100 (I am completely confident that the perpetrator was not in the lineup). The anchors were rewritten for sequential lineups to reflect the fact that the retrospective judgments were made for each individual photograph.

4. Lineup composition was adjusted. In Experiment 2, high-similarity foil 1 in the innocent-suspect lineup was rated as slightly more similar to the perpetrator ($m = 2.20$) than the innocent suspect was to the perpetrator ($m = 2.02$). Although this situation is not impossible, it should

occur infrequently when foils are selected based on their similarity to the suspect (Clark & Tunnicliff, 2001; Navon, 1992). In Experiment 2, high-similarity foil 1 was replaced with a different high-similarity foil with a lower similarity rating to the perpetrator ($m = 1.72$). The corresponding high-similarity foil in the guilty suspect lineup was replaced as well to ensure the similarity between each foil and the suspect was equated across guilty and innocent suspect lineups. The design and composition of the lineups are illustrated in Table 11. The lineups appear in Figure 11. This change yielded lineups with lower similarity high-similarity lineup members on average in Experiment 3 (Guilty Suspect lineups $m = 3.165$ and Innocent Suspect lineups $m = 3.105$) compared to Experiment 2 (Guilty Suspect lineups $m = 3.345$ and Innocent Suspect lineups $m = 3.360$). This change could also have been warranted based on results by Fitzgerald, Oriet, and Price (2015). Using morph faces to control similarity relations between the suspect and foils, Fitzgerald et al. found that moderately high similarity lineups produced higher correct identification rates with little increase in false identification rates compared to very high similarity lineups, indicating that foils matched to the suspect could be too similar in some cases. The high-similarity neighbor advantage in Experiment 2 may have been small because the high-similarity neighbors were too similar to the guilty suspect.

5. The delay task between the staged crime and lineup differed from the tasks in Experiments 1 and 2. In Experiment 3, participants performed a four-part Stroop task (Stroop, 1935). Each part lasted 5 minutes for a total duration of 20 minutes.

Table 11

Mean similarities of lineup members to guilty and innocent suspects. Separate participants in the similarity rating study rated the similarity of the guilty suspect to the innocent suspect (To S_i) and the similarity of the innocent suspect to the guilty suspect (To S_g) as well.

	Guilty Suspect	Innocent Suspect	
Member		To I_s	To G_s
Suspect	–	3.13	2.02
HS-Foil 1	3.22	3.16	2.00
HS-Foil 2	3.11	3.05	1.72
LS-Foil 1	1.27	1.27	1.10
LS-Foil 2	1.37	1.40	1.19

Figure 11. Guilty-suspect and innocent-suspect high-similarity lineups from Experiment 3. The suspect appears in position 3. High-similarity foil 1 from Experiment 2 was replaced with high-similarity foil 2 (position 4) in this figure for both guilty and innocent suspect lineups.

Results

As in the previous Experiments, three sets of analyses are presented. The first set presents ROC curves and pAUC comparisons that examined differences in discriminability by neighborhood, lineup presentation, and suspect position. The second set of analyses assessed identification responses for neighborhood and lineup presentation conditions separately for guilty and innocent suspect lineups using log linear analyses. The third set of analyses investigated the relationship between confidence and accuracy using confidence-accuracy calibration curves.

Discriminability

The first set of analyses evaluated the accuracy of witnesses as a function of lineup presentation, neighborhood similarity and suspect position using d' and pAUC. Simultaneous lineups yielded significantly higher accuracy (d' = 0.868 and pAUC = .187) compared to sequential lineups (d' = .588 and pAUC = .147), based on d', $G = 2.129$, $p = .033$ and pAUC, $D = 2.694$, $p = .007$. Table 12 shows d' and pAUC analyses comparing simultaneous to sequential lineups within high and low-similarity neighborhood conditions by suspect position. The corresponding ROC curves for high- and low-similarity neighbor conditions and simultaneous and sequential lineups appear in Figure 12.

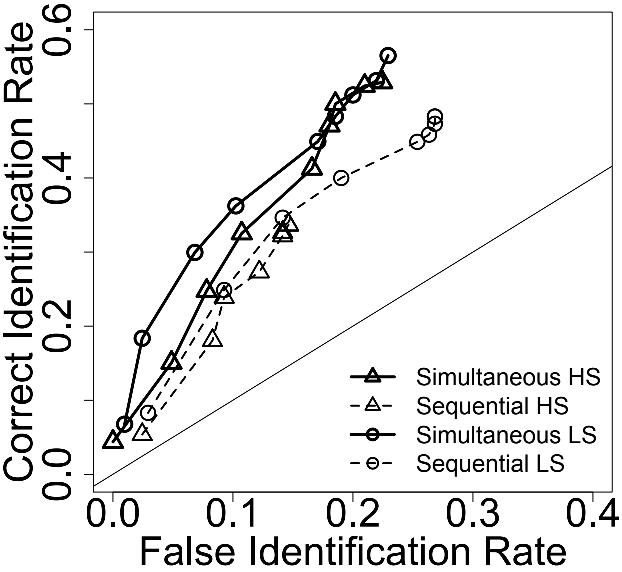


Figure 12. Receiver Operating Characteristic (ROC) plot for high- and low-similarity neighbor simultaneous and sequential conditions. The diagonal line represents chance performance.

Table 12

Discriminability analyses by lineup presentation, neighborhood, and suspect position.

Lineup		Sequential		Simultaneous		d' analyses		pAUC analyses	
Presentation	Position	d'	pAUC	d'	pAUC	G	p	D	p
High	Overall	0.632	0.141	0.844	0.179	1.105	.269	1.910	.056
Similarity	2	0.539	0.127	1.267	0.219	2.158	.031	2.627	.009
Neighbors	3	0.603	0.134	0.729	0.166	0.375	.707	0.941	.347
	4	0.750	0.165	0.522	0.153	−0.692	.489	−0.366	.714
Low	Overall	0.577	0.155	0.923	0.194	1.888	.059	1.861	.063
Similarity	2	0.669	0.161	1.271	0.229	1.879	.060	1.856	.064
Neighbors	3	0.621	0.172	0.802	0.194	0.576	.565	0.582	.561
	4	0.456	0.129	0.670	0.162	0.654	.513	0.955	.340

Within the high-similarity neighbors condition, the pAUC was marginally larger for simultaneous lineups (.179) than for sequential lineups (.141), $D = 1.910$, $p = .056$, owing to a large accuracy advantage of simultaneous lineups (.219) compared to sequential lineups (.127) in suspect position 2, $D = 2.627$, $p = .009$ for high-similarity neighbors. Accuracy was only slightly higher for simultaneous (.166) compared to sequential lineups (.134) in suspect position 3, $D = 0.941$, $p = .347$, and lower for simultaneous lineups (.522) compared to sequential lineups (.750) in suspect position 4, $D = -0.366$, $p = .714$.

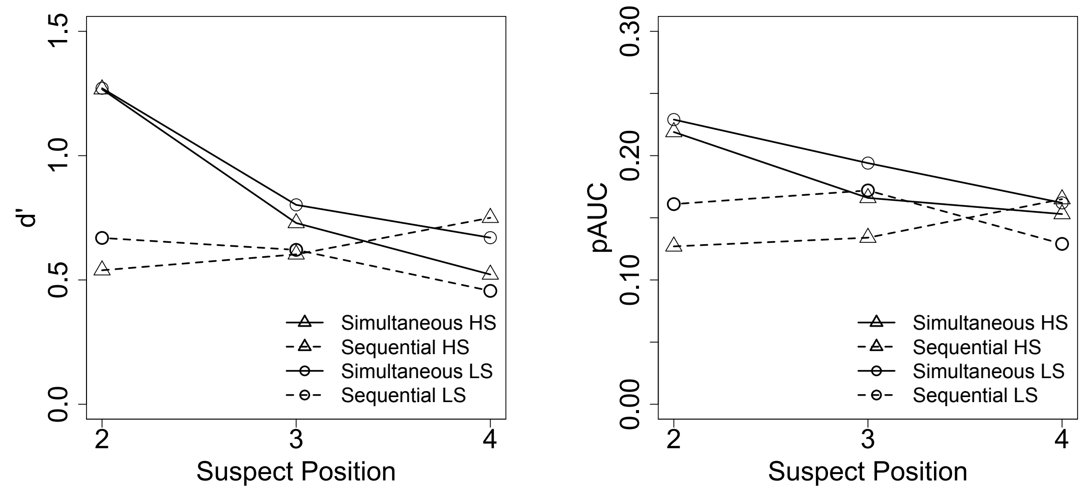


Figure 13. Plots of d' and pAUC by lineup type, neighborhood, and suspect position.

Within the low-similarity neighbors condition, pAUC was marginally higher overall for simultaneous (.194) compared to sequential lineups (.155) , $D = 1.861$, $p = .063$. This difference was also due to a marginal advantage for simultaneous (.229) compared to sequential lineups (.161) in suspect position 2, $D = 1.856$, $p = .064$. Accuracy was only slightly higher for simultaneous compared to sequential lineups in suspect positions 3 (sim = .194 and seq = .172) and 4 (sim = .162 and seq = .129), $D = 0.582$, $p = .561$, and $D = 0.955$, $p = .340$, respectively.

The pattern was the same collapsing over neighborhood—simultaneous lineups yielded significantly higher accuracy based on pAUC compared to sequential lineups in suspect position 2 (sim = .225 and seq = .143, $D = 3.202$, $p = .001$), but not 3 (sim = .179 and seq = .150, $D = 1.131$, $p = .258$), or 4 (sim = .157 and seq = .148, $D = 0.398$, $p = .690$). Accuracy for simultaneous lineups declined across the lineup, (Position 2 compared to 3: $D = 1.747$, $p = .081$; Position 2 compared to 4: $D = 2.639$, $p = .008$; Position 3 compared to 4: $D = 0.832$, $p = .405$), but accuracy for sequential lineups varied very little ($.787 < p < .934$).

The ROC analyses showed that the advantage of simultaneous lineups was largely due to an accuracy advantage in suspect position 2. The accuracy advantage for simultaneous lineups was less stable across positions, especially in suspect position 4. In suspect position 4, sequential lineups yielded a small advantage when a high-similarity foil preceded the suspect. The suspect position effects on accuracy are illustrated in Figure 13, which shows d' and pAUC values across suspect position for simultaneous and sequential high and low-similarity neighbor lineups.

The simultaneous advantage was more pronounced in the high-similarity neighbors condition, although high-similarity neighbors did not produce the same accuracy advantage as observed in Experiment 2. In fact, for simultaneous lineups, accuracy was somewhat (but not significantly) lower for high-similarity neighbors (.179) compared to low-similarity neighbors (.194) in Experiment 3, $D = -0.695$, $p = .487$. Why did Experiment 3 fail to replicate the high-similarity advantage found in Experiment 2 for simultaneous lineups? One argument pertains to the similarity of the high-similarity lineup members. As mentioned, the highest-similarity lineup member in Experiment 2 was replaced by a moderately similar lineup member in Experiment 3, lowering the overall neighborhood similarity. The comparisons between the suspect and the moderate similarity lineup member may have resulted in fewer shared features to ignore, and fewer diagnostic features.

Identification Responses

Table 13 and Figure 14 show the response proportions for suspect, foil, and non-identification responses for high-similarity neighbor and low-similarity neighbor conditions, as a function of suspect guilt, suspect position, and lineup presentation.

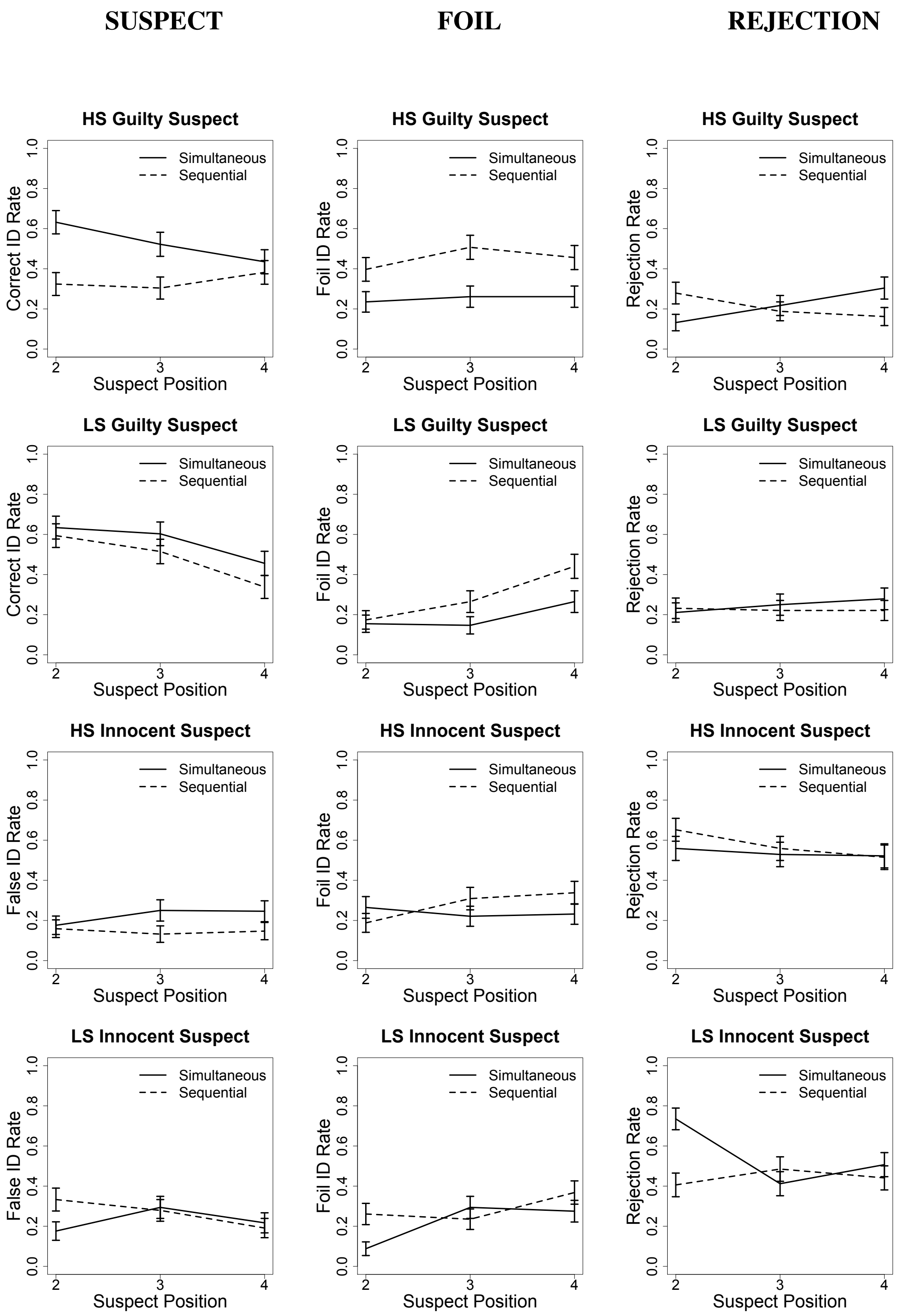


Figure 14. Identification responses plotted by lineup type, lineup presentation, neighborhood, and suspect position. Error bars represent standard error of a proportion.

Table 13

Frequencies and proportions of suspect, foil, and no pick responses for neighborhood, lineup presentation, suspect guilt or innocence, and suspect position. P-value (p) for 2 × 2 chi-square comparisons.

			Guilty Suspect			Innocent Suspect		
Lineup Presentation	Suspect Position	Response	Sequential f (%)	Simultaneous f (%)	*p*	Sequential f (%)	Simultaneous f (%)	*p*
High Similarity	Collapsed	Suspect	69 (.336)	109 (.529)	< .001*	30 (.146)	46 (.224)	.042*
		Foil	93 (.454)	52 (.252)	< .001*	57 (.278)	49 (.239)	.367
		No Pick	43 (.210)	45 (.218)	.830	118 (.576)	110 (.463)	.427
	2	Suspect	22 (.324)	43 (.632)	< .001*	11 (.159)	12 (.176)	.790
		Foil	27 (.397)	16 (.235)	.043*	13 (.188)	18 (.265)	.286
		No Pick	19 (.279)	9 (.132)	.034*	45 (.652)	38 (.559)	.264
	3	Suspect	21 (.304)	36 (.522)	.010*	9 (.132)	17 (.250)	.081
		Foil	35 (.507)	18 (.261)	.003*	21 (.309)	15 (.221)	.244
		No Pick	13 (.188)	15 (.217)	.672	38 (.559)	36 (.529)	.731
	4	Suspect	26 (.382)	30 (.435)	.533	10 (.147)	17 (.246)	.144
		Foil	31 (.455)	18 (.261)	.017*	23 (.338)	16 (.232)	.168
		No Pick	11 (.162)	21 (.304)	.049*	35 (.515)	36 (.522)	.934
Low Similarity	Collapsed	Suspect	99 (.483)	117 (.565)	.094	55 (.268)	47 (.229)	.361
		Foil	60 (.293)	39 (.188)	.013*	59 (.288)	45 (.220)	.112
		No Pick	46 (.224)	51 (.246)	.599	91 (.444)	113 (.551)	.030*
	2	Suspect	41 (.594)	45 (.633)	.630	23 (.333)	12 (.176)	035*
		Foil	12 (.174)	11 (.155)	.762	18 (.261)	6 (.090)	.008*
		No Pick	16 (.232)	15 (.211)	.769	28 (.406)	50 (.735)	< .001*
	3	Suspect	35 (.515)	41 (.603)	.300	19 (.279)	20 (.294)	.850
		Foil	18 (.265)	10 (.147)	.090	16 (.235)	20 (.294)	.437
		No Pick	15 (.221)	17 (.250)	.686	33 (.485)	28 (.412)	.389
	4	Suspect	23 (.338)	31 (.456)	.161	13 (.191)	15 (.217)	.704
		Foil	30 (.441)	18 (.265)	.031*	25 (.368)	19 (.275)	.247
		No Pick	15 (.221)	19 (.279)	.428	30 (.441)	35 (507)	.439

Separate log-linear analyses were performed for guilty and innocent-suspect lineups for each response type with neighborhood, lineup presentation, and suspect position entered as predictors. The results of that analysis are displayed in Table 14. Table 15 depicts additional log-linear analysis results broken up by lineup presentation and suspect guilt and innocence. Those analysed examine the contribution of identification rates to discriminability.

Guilty Suspect lineups

Correct identifications. In the full log linear model including lineup presentation, neighborhood, and suspect position as predictors, simultaneous lineups yielded more correct identifications

Table 14

Log linear analyses for the full model including response (suspect identifications, foil identifications or rejections), neighborhood, presentation, and suspect position for guilty and innocent suspect lineups. Statistically significant results ($p < .05$*) appear in bold.*

		Guilty Suspect		Innocent Suspect	
Response	Model Terms	Deviance	p	Deviance	p
Suspect ID	Neighborhood × Presentation × Suspect Position	3.952	.139	0.321	.852
	Neighborhood × Suspect Position	4.222	.121	1.483	.477
	Presentation × Suspect Position	1.037	.596	4.926	**.085**
	Neighborhood × Presentation	2.770	**.096**	4.541	**.033**
	Suspect Position	11.717	**.003**	1.246	.536
	Presentation	15.737	< **.001**	0.462	.497
	Neighborhood	6.865	**.009**	4.865	**.027**
Foil ID	Neighborhood × Presentation × Suspect Position	0.476	.788	9.164	**.010**
	Neighborhood × Suspect Position	6.875	**.032**	1.544	.462
	Presentation × Suspect Position	1.206	.547	1.087	.581
	Neighborhood × Presentation	1.079	.299	0.245	.620
	Suspect Position	9.136	**.010**	7.881	**.019**
	Presentation	23.552	< **.001**	3.103	**.078**
	Neighborhood	12.542	< **.001**	0.026	.873
Lineup Rejection	Neighborhood × Presentation × Suspect Position	2.533	.282	8.735	**.013**
	Neighborhood × Suspect Position	0.070	.966	0.589	.745
	Presentation × Suspect Position	6.728	**.035**	4.009	.135
	Neighborhood × Presentation	0.046	.830	4.435	**.035**
	Suspect Position	0.699	.705	6.173	**.046**
	Presentation	0.278	.598	0.962	.327
	Neighborhood	0.537	.464	2.820	**.093**

(.547) compared to sequential lineups (.409), $\chi^2(1, N = 823) = 15.737$, $p < .001$, $r = .138$, and low-similarity neighbors yielded more correct identifications (.524) compared to high-similarity neighbors (.433), $\chi^2(1, N = 823) = 6.865$, $p = .009$, $r = .091$, collapsing over lineup presentation. These main effects were qualified by a marginally significant interaction between neighborhood and lineup presentation, which indicated that patterns of correct identifications for simultaneous and sequential lineups differed depending on the neighborhood of the lineup, $\chi^2(1, N = 823) = 2.770$, $p = .096$, $r = .058$. The following comparisons involving simultaneous and sequential presentation are reported separately for high and low-similarity neighbor lineups.

For high-similarity neighbor lineups, the correct identification rate was significantly higher for simultaneous (.529) compared to sequential lineups (.337), $\chi^2(1, N = 411) = 15.622$, $p < .001$, $r = .195$. This effect varied marginally by the position of the suspect, $\chi^2(2, N = 411) = 4.683$, $p = .096$. The correct identification rate was significantly higher for simultaneous lineups compared to sequential lineups in suspect position 2 (sim = .632, seq = .324) and 3 (sim = .522, seq = .304)

Table 15

Log linear analyses for high and low-similarity neighborhoods including response (suspect identifications, foil identifications or rejections), presentation, and suspect position for guilty and innocent suspect lineups. Statistically significant results ($p < .05$) appear in bold.

			Guilty Suspect		Innocent Suspect	
Neighborhood	Response	Model Terms	Deviance	p	Deviance	p
High Similarity	Suspect ID	Presentation × Suspect Position	4.683	**.096**	1.162	.560
		Suspect Position	1.751	.417	0.426	.808
		Presentation	15.622	< **.001**	4.161	**.041**
	Foil ID	Presentation × Suspect Position	0.363	.834	3.588	.166
		Suspect Position	1.480	.477	1.285	.526
		Presentation	18.410	< **.001**	0.815	.367
	Lineup Rejection	Presentation × Suspect Position	8.656	**.013**	0.755	.686
		Suspect Position	0.461	.794	2.246	.325
		Presentation	0.046	.830	0.633	.426
Low Similarity	Suspect ID	Presentation × Suspect Position	0.441	.802	3.842	.147
		Suspect Position	14.139	**.001**	2.554	.279
		Presentation	2.800	**.094**	0.836	.361
	Foil ID	Presentation × Suspect Position	1.364	.505	6.655	**.036**
		Suspect Position	14.535	< **.001**	8.150	**.017**
		Presentation	6.171	**.013**	2.531	.112
	Lineup Rejection	Presentation × Suspect Position	0.598	.742	12.01	**.002**
		Suspect Position	0.316	.854	4.529	.104
		Presentation	0.277	.599	4.731	**.030**

but not 4 (sim = .435, seq = .382), $\chi^2(1, N = 136) = 12.996$, $p = .0003$, $r = .309$, $\chi^2(1, N = 138) = 6.725$, $p = .010$, $r = .221$ and $\chi^2(1, N = 137) = 0.390$, $p = .533$, $r = .053$, respectively. This result provided support for the prediction that high-similarity neighbor lineups would not facilitate performance in sequential lineups. The impact of neighborhood, however, seemed to “wash out" by suspect position 4 for simultaneous lineups.

For low-similarity neighbor lineups, correct identifications were marginally higher for simultaneous lineups (.565) compared to sequential lineups (.483), $\chi^2(1, N = 412) = 2.800$, $p = .094$, $r = .082$. Relative to the high-similarity neighbor results above, when the suspect was surrounded by low-similarity neighbors, the correct identification did not increase as much for simultaneous compared to sequential lineups; thus, surrounding the suspect with low-similarity neighbors reduced the simultaneous advantage. The correct identification rate was slightly higher for simultaneous lineups as well (.565 compared to .529), which was at odds with the high-similarity correct identification advantage found in Experiment 2.

Within low-similarity neighbor lineups, suspect position significantly impacted correct identifications, independent of lineup presentation, $\chi^2(2,412) = 14.139$, $p = .001$. Correct identifications were significantly higher when the suspect appeared in position 2 (.614) compared to 4 (.397), $\chi^2(1,N = 276) = 13.024$, $p = .0003$, $r = .217$, significantly higher when the suspect appeared in position 3 (.559) compared to 4, $\chi^2(1,N = 272) = 7.132$, $p = .008$, $r = .162$, and nonsignificantly higher when the suspect appeared in position 2 compared to 3, $\chi^2(1,N = 276) = 0.875$, $p = .350$, $r = .056$. In other words, the correct identification rate declined across the lineup, from left to right. The same pattern of a decline in correct identification rates across suspect position holds combining the high and low-similarity neighbor lineups in the full model, $\chi^2(2,N = 823) = 11.717$, $p = .003$. Correct identifications collapsing over lineup presentation were significantly higher when the suspect appeared in position 2 (.547) compared to 4 (.403), $\chi^2(1,N = 549) = 11.438$, $p = .0007$, $r = .144$, marginally higher when the suspect appeared in position 3 (.485) compared to 4, $\chi^2(1,N = 547) = 3.767$, $p = .052$, $r = .082$, and nonsignificantly higher when the suspect appeared in position 2 compared to 3, $\chi^2(1,N = 550) = 2.096$, $p = .148$, $r = .062$. Thus, there was a consistent decline in correct identifications as the lineup progressed that was somewhat more pronounced for low-similarity neighbors.

The neighborhood effect was also examined by comparing high- and low-similarity neighborhoods separately for sequential and simultaneous lineups. For sequential lineups, the effect of neighborhood varied by the position of the suspect, $\chi^2(2,N = 410) = 7.700$, $p = .021$. The correct identification rate was higher when the guilty suspect was preceded by a low-similarity foil than when preceded by a high-similarity foil in suspect positions 2 and 3, $\chi^2(1,N = 137) = 10.102$, $p = .001$, $r = .272$; $\chi^2(1,N = 137) = 6.271$, $p = .012$, $r = .214$; but not 4, $\chi^2(1,N = 136) = 0.287$, $p = .592$, $r = .046$, respectively. Thus, there was a large correct identification rate advantage for low-similarity neighbor sequential lineups in suspect positions 2 and 3 but not 4. For simultaneous lineups, correct identifications were slightly but not significantly higher for the low-similarity neighbor condition (.565) compared to the high-similarity neighbor condition (.529), $\chi^2(1,N = 413) = 0.543$, $p = .461$, $r = .036$.

Foil identifications. In the full model, the foil identification rate was significantly higher for sequential (.372) compared to simultaneous lineups (.220), $\chi^2(1,N = 823) = 23.552$, $p < .001$, $r = .169$. When witnesses cannot see all lineup members at once, they may be more likely to “spend" their identification choice on the foils. The effect of lineup presentation on foil identifications did not depend on neighborhood, $\chi^2(1, N = 823) = 1.079$, $p = .299$, $r = .036$. The effect of neighborhood on the foil identification rate, however, did depend on the position of the suspect, $\chi^2(2, N = 823) = 6.875$, $p = .032$. Specifically, the foil identification rate varied depending on suspect position for low-similarity neighbors, $\chi^2(2, N = 412) = 14.535$, $p = .0007$, but not for high-similarity neighbors, $\chi^2(2, N = 411) = 1.480$, $p = .477$. For low-similarity neighbor lineups, the foil identification rate was significantly lower when the suspect appeared in position 2 (.164) compared to 4 (.353), $\chi^2(1, N = 276) = 12.850$, $p = .0003$, $r = .216$, and 3 (.206) compared to 4, $\chi^2(1, N = 272) = 7.304$, $p = .007$, $r = .164$, but not 2 compared to 3, $\chi^2(1, N = 276) = 0.792$, $p = .373$, $r = .054$.

Lineup rejections. Simultaneous lineups yielded nonsignificantly more incorrect rejections (.232) compared to sequential lineups (.217), $\chi^2(1, N = 823) = 0.278$, $p = .598$, $r = .018$, although the effect of lineup presentation on the incorrect rejection rate depended on suspect position, $\chi^2(2, N = 823) = 6.728$, $p = .035$. When the suspect appeared in position 2, incorrect rejections were marginally higher for sequential lineups (.255) compared to simultaneous lineups (.173), $\chi^2(1, N = 276) = 2.815$, $p = .093$, $r = .101$. When the suspect appeared in position 3 or 4, however, incorrect rejections were somewhat higher for simultaneous lineups (pos 3 = .233, pos 4 = .292) compared to sequential lineups (pos = 3 = .204, pos 4 = .191), $\chi^2(1, N = 274) = 0.341$, $p = .559$, $r = .035$, and $\chi^2(1, N = 273) = 3.783$, $p = .052$, $r = .118$, respectively. In sequential lineups, witnesses seem to wait until later suspect positions to make an identification, resulting in more miss errors in position 2.

Neighborhood had no reliable impact on the incorrect rejection rate, although for simultaneous and sequential lineups, witnesses made nonsignificantly fewer incorrect rejections when the suspect was surrounded by high-similarity lineup members (sim = .218, seq = .210) compared to low-similarity lineup members (sim = .246, seq = .224), $\chi^2(1, N = 413) = 0.452$, $p = .502$, $r = .033$ and $\chi^2(1, N = 410) = 0.129$, $p = .719$, $r = .018$, respectively.

Innocent suspect lineups

False identifications. In the full log linear model, false identifications were slightly but not significantly higher in simultaneous (.227) compared to sequential lineups (.207), $\chi^2(1, N = 820) = 0.462$, $p = .497$, $r = .024$. The effect of lineup presentation on false identifications differed depending on neighborhood, $\chi^2(1, N = 820) = 4.541$, $p = .033$, $r = .074$. For high-similarity neighbor lineups, the false identification rate was higher when the lineup was presented simultaneously (.224) than when the lineup was presented sequentially (.146), $\chi^2(1, N = 410) = 4.161$, $p = .042$, $r = .100$. For low-similarity neighbor lineups, the false identification rate was slightly (nonsignificantly) higher when the lineup was presented sequentially (.268) than when the lineup was presented simultaneously (.229), $\chi^2(1, N = 410) = 0.836$, $p = .361$, $r = .045$.

The interaction between lineup presentation and neighborhood can also be considered by comparing high- and low-similarity neighborhoods by lineup presentation method. For sequential lineups, the false identification rate was higher when the guilty suspect was preceded by a low-similarity foil (.268) than when preceded by a high-similarity foil (.146), $\chi^2(1, N = 410) = 9.388$, $p = .002$, $r = .151$. For simultaneous lineups, the false identification rate was slightly (nonsignificantly) higher when the guilty suspect was surrounded by low-similarity foils (.229) than when surrounded by high-similarity foils (.224), $\chi^2(1, N = 410) = 0.014$, $p = .906$, $r = .009$. In other words, low-similarity neighbors increased both correct and false identifications for sequential lineups and (to a lesser extent) simultaneous lineups.

In the full log linear model, the effect of lineup presentation on false identifications differed marginally depending on the position of the suspect, $\chi^2(2, N = 820) = 4.926$, $p = .085$. False identifications were slightly but not significantly higher when the suspect appeared in position 2 for sequential lineups (.246) compared to simultaneous (.176) lineups, $\chi^2(1, N = 274) = 2.006$, $p = .157$, $r = .086$. In contrast, false identifications were slightly but not significantly higher when the suspect appeared in positions 3 and 4 for simultaneous (pos 3 = .272. , pos 4 =.232) compared to sequential lineups, (pos 3 = .206. , pos 4 =.169), $\chi^2(1, N = 272) = 1.638$, $p = .201$, $r = .078$, and $\chi^2(1, N = 274) = 1.682$, $p = .195$, $r = .078$, respectively. Thus, simultaneous lineups generally increased false identifications, though depending on the position of the suspect, sequential lineups too can increase the false identification rate.

Foil identifications. In innocent-suspect lineups, the foil identification rate was marginally higher for sequential lineups (.283) compared to simultaneous lineups (.229), $\chi^2(1, N = 820) = 3.103$, $p = .078$, $r = .062$, although this result was qualified by a significant interaction between neighborhood, lineup presentation, and suspect position, $\chi^2(2, N = 820) = 9.164$, $p = .010$. This interaction is largely driven by an interaction between lineup presentation and suspect position in low-similarity neighbor lineups, $\chi^2(2, N = 410) = 6.655$, $p = .036$. This interaction was not significant in high-similarity lineups, $\chi^2(2, N = 410) = 3.588$, $p = .166$. In low-similarity lineups, when the suspect appeared in position 2, simultaneous lineups produced significantly fewer foil identifications (.088) compared to sequential lineups (.260), $\chi^2(1, N = 137) = 7.064$, $p = .008$, $r = .227$. When the suspect appeared in positions 3 and 4, foil identifications did not differ reliably for simultaneous and sequential lineups, $\chi^2(1, N = 136) = 0.604$, $p = .437$, $r = .067$, and $\chi^2(1, N = 137) = 1.338$, $p = .247$, $r = .099$, respectively.

Lineup rejections. Sequential lineups produced nonsignificantly fewer correct rejections (.510) compared to simultaneous lineups (.544), $\chi^2(1, N = 820) = 0.962$, $p = .327$, $r = .034$. The highest order interaction (neighborhood × presentation × suspect position) in the full log linear model was statistically significant $\chi^2(2, N = 820) = 8.735$, $p = .013$. This interaction was driven by an interaction between presentation and suspect position for low-similarity, $\chi^2(2, N = 410) = 12.010$, $p = .002$, but not high-similarity lineups, $\chi^2(2, N = 410) = 0.755$, $p = .686$. In low-similarity lineups, when the suspect was in position 2, correct rejections were higher in simultaneous lineups (.735) compared to sequential lineups (.405), $\chi^2(1, N = 137) = 15.165$, $p < .0001$, $r = .333$. In contrast, correct rejections did not differ reliably between simultaneous and sequential lineups in positions 3 and 4, $\chi^2(1, N = 136) = 0.743$, $p = .389$, $r = .074$ and $\chi^2(1, N = 137) = 0.600$, $p = .439$, $r = .066$.

Confidence and Accuracy

The point-biserial correlations between confidence and accuracy for high-similarity neighbor simultaneous lineups, $r = .212$ ($N = 202$), $p = .002$, and high-similarity neighbor sequential lineups, $r = .157$ ($N = 154$), $p = .050$ did not differ statistically from each other, $z = 0.530$, $p = .596$. The point-biserial correlations for low-similarity neighbor simultaneous lineups, $r = .242$ ($N = 207$),

$p = .0004$, and low-similarity neighbor sequential lineups, $r = .238$ ($N = 211$), $p = .0005$ did not differ statistically from each other, $z = 0.484$, $p = .969$. The suspect identification point-biserial correlations for high-similarity neighbor simultaneous lineups, $r = .175$ ($N = 153$), $p = .029$, and high-similarity neighbor sequential lineups, $r = -.004$ ($N = 97$), $p = .967$ did not differ statistically from each other, $z = 1.137$, $p = .171$. Just considering suspect identifications, the point-biserial correlations for low-similarity neighbor simultaneous lineups, $r = .174$ ($N = 162$), $p = .026$, and low-similarity neighbor sequential lineups, $r = .143$ ($N = 152$), $p = .075$ did not differ statistically from each other, $z = 0.280$, $p = .780$.

For both simultaneous and sequential lineups, the confidence-accuracy calibration curves are shown in Figure 15. Confidence was partitioned into the following bins: 0-30, 31-50, 51-70, 71-80, 81-99, and 100. For simultaneous lineups, both high- and low-similarity neighbor lineups showed a correspondence between confidence and accuracy in terms of the calibration curves and correlations between confidence and accuracy: high-similarity neighbors, $r = .928$, $p = .008$, and low-similarity neighbors, $r = .886$, $p = .019$. Unlike in Experiment 2, the high-similarity condition did not attenuate the relationship between confidence and accuracy.

Both the high- and low-similarity neighborhood curves for simultaneous lineups are generally higher than those for sequential lineups, indicating consistently higher accuracy across most levels of confidence. Confidence and accuracy were relatively poorly calibrated for sequential lineups. For high-similarity neighbor sequential lineups, confidence and accuracy mirrored performance in simultaneous lineups for confidence levels below 71-80, but dropped off at the higher levels of confidence, $r = .274$, $p = .599$. The relationship between confidence and accuracy was concave for low-similarity neighbor sequential lineups, $r = .242$, $p = .644$, with highest accuracy at the lowest and highest levels of confidence.

The sequential calibration curves may have been unstable because of lower numbers of responses at certain levels of confidence. For example, there were only one false identification and five correct identifications made in the 0-30 confidence interval for low-similarity neighbor sequential lineups, which is why the calibration curve in Figure 15 had a concave shape. The five correct identifications were evenly distributed across suspect position (correct: pos 2 = 2, pos 3 = 1, and pos 4 = 2; false: pos 4 = 1). A chi-square test comparing correct and false identifications made with confidence

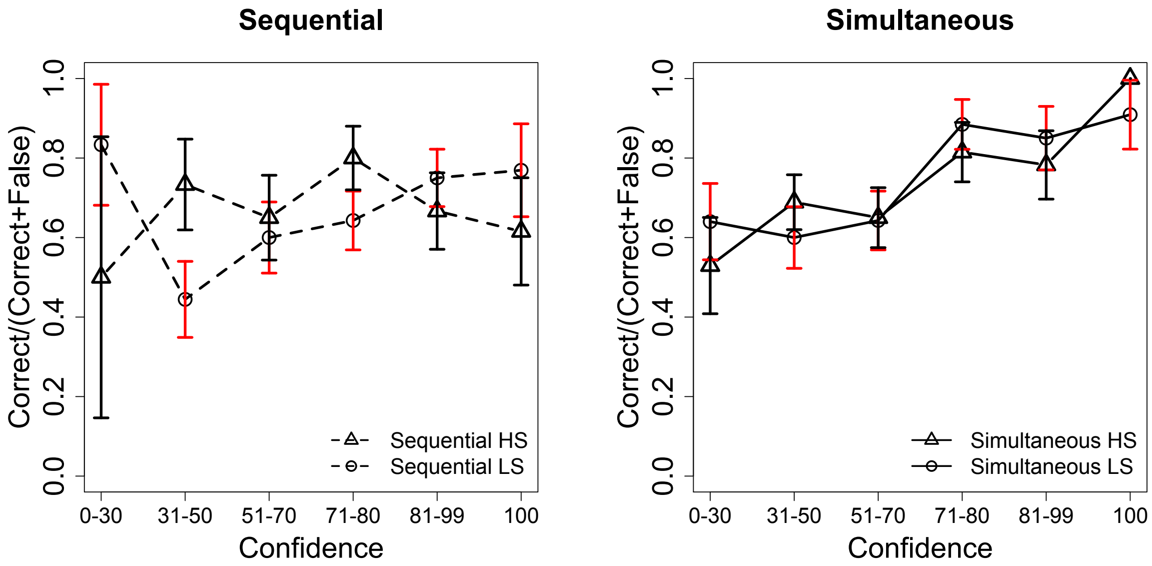


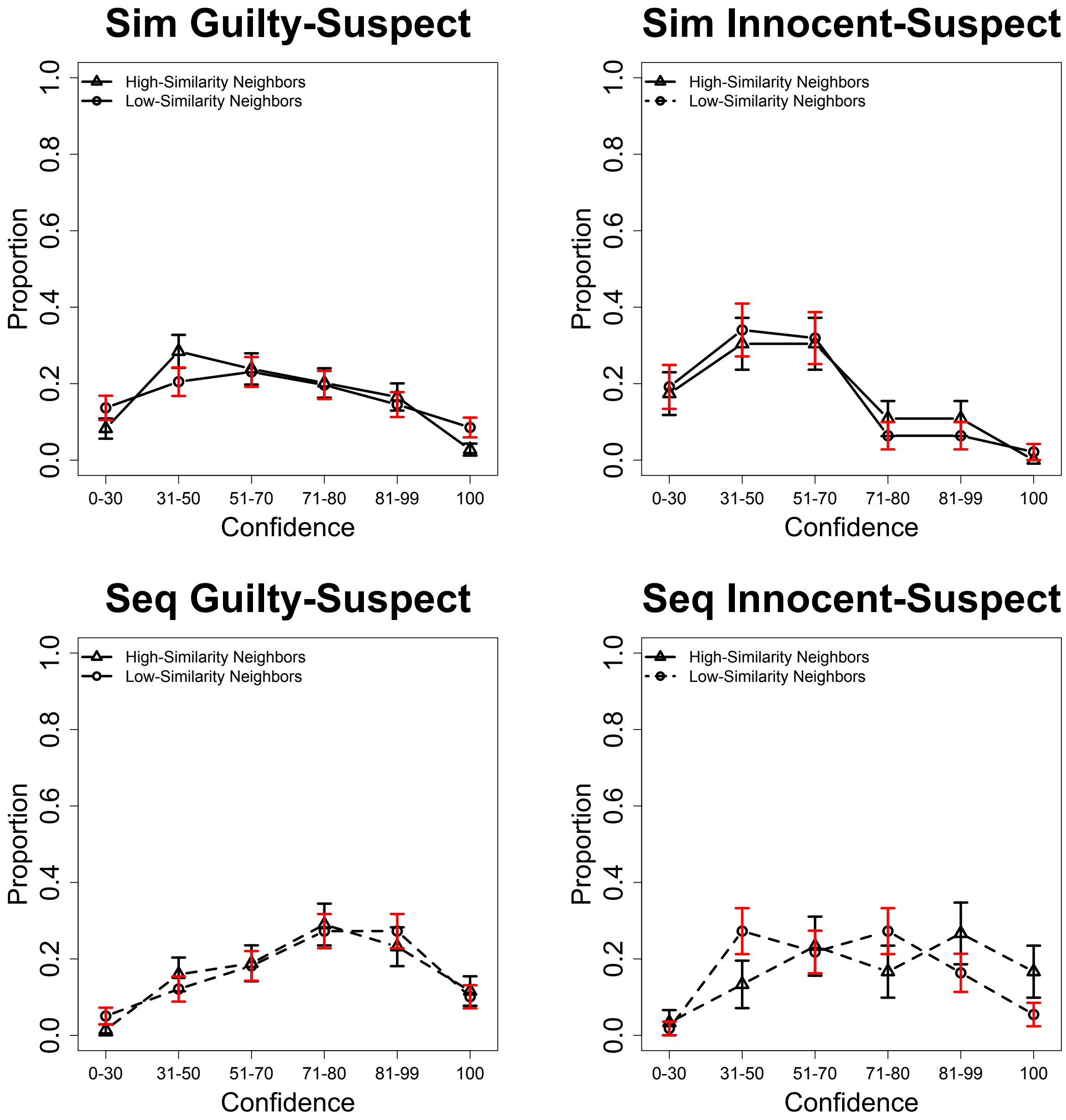
Figure 15. *Confidence accuracy calibration curves for sequential and simultaneous high and low-similarity neighbor lineups. Error bars represent standard error of a proportion. Low-similarity neighbor error bars are red, high-similarity neighbor error bars are black.*

of 1 for high and low-similarity neighbors was not statistically significant, $\chi^2(1, N = 8) = 0.889$, $p = .346$, $r = .333$, indicating that the difference of appearance of the curves for high- versus low-similarity neighbors may be spurious. In addition, there were similar numbers of correct (8) and false (5) identifications for the high-similarity neighbor sequential lineups at the highest level of confidence, yielding the drop in accuracy at the highest level of confidence.

The distributions of confidence for suspect identifications from guilty and innocent-suspect lineups are shown in Figure 16. The distributions of confidence for the high-similarity neighbor and low-similarity neighbor simultaneous and sequential conditions were not significantly different for guilty-suspect lineups, simultaneous: $\chi^2(5, N = 226) = 6.415$, $p = .268$, sequential: $\chi^2(5, N = 168) = 2.312$, $p = .805$, nor for innocent-suspect lineups, simultaneous: $\chi^2(5, N = 93) = 2.216$, $p = .819$, sequential: $\chi^2(5, N = 85) = 6.448$, $p = .265$.

Discussion

Experiment 3 examined whether the simultaneous advantage depended on comparisons between lineup members. Lineups were presented simultaneously or sequentially, with high or low-


Sim Guilty-Suspect

Sim Innocent-Suspect

Seq Guilty-Suspect

Seq Innocent-Suspect

Figure 16. Distributions of confidence. Sim = Simultaneous (top), Seq = Sequential (bottom)

similarity neighbors. In simultaneous lineups, the high-similarity neighbor condition surrounded the suspect with high-similarity foils, facilitating comparisons between the best matching lineup members; whereas the low-similarity neighbor condition surrounded the suspect with low-similarity foils, making direct comparisons between the best matching lineup members more difficult. In sequential lineups, the high-similarity condition was instantiated by placing a high-similarity foil directly before and directly after the suspect. The low-similarity sequential condition placed a low-similarity foil directly before and directly after the suspect. To the extent that comparative processes underlie the simultaneous advantage, the high-similarity neighbor condition should improve identification performance more for simultaneous lineups and less for sequential lineups.

There were two main results in the present study. First, in concert with recent studies showing higher accuracy for simultaneous lineups compared to sequential lineups (Dobolyi & Dodson, 2013; Gronlund, Carlson, et al., 2013; Mickes et al., 2012), the discriminability results from the present study showed that simultaneous lineups were more accurate compared to sequential lineups collapsing over neighborhood, especially when the suspect appeared in position 2. The simultaneous advantage derived from an increase in correct identifications, a slight increase in correct rejections, and fewer foil identifications in guilty suspect simultaneous lineups compared to sequential lineups. Although simultaneous lineups were consistently more accurate than sequential lineups across different combinations of conditions, the simultaneous advantage decreased across suspect position. Discriminability improved somewhat for sequential lineups in the later suspect positions, consistent with previous research reporting facilitation of performance in late sequential positions (Gronlund et al., 2009).

The central result of Experiment 3, though, concerned the impact of neighborhood and suspect position on the relationship between lineup presentation and accuracy. Witnesses were more accurate when the high-similarity neighbor lineup was presented simultaneously compared to sequentially when the suspect appeared in position 2. This discriminability difference was due to a large increase in correct identifications and a very small increase in false identifications for simultaneous compared to sequential lineups. The accuracy advantage of high-similarity simultaneous lineups declined across levels of suspect position, becoming a slight sequential advantage in suspect position 4 (the only time sequential lineup performance exceeded simultaneous performance), as the high-similarity simultaneous correct identification rate advantage shrank and the false identification rate disadvantage increased relative to suspect position 2. Further, discriminability in sequential lineups increased across the lineup, as witnesses tended to save their identification choice until later in the lineup. Witnesses may also have learned about the distribution of features over the course of the sequential lineup, allowing them to make comparisons based on memory in later suspect positions, thereby improving accuracy relative to position 2.

Witnesses were also slightly more accurate when the low-similarity neighbor lineup was presented simultaneously compared to sequentially in suspect position 2, although this result was due to a decrease in the false identification rate for simultaneous lineups relative to sequential lineups, rather than a correct identification rate advantage as was the case in the high-similarity neighbors condition. The low-similarity simultaneous advantage was far less pronounced than the high-similarity simultaneous advantage because the correct identification rate was only slightly higher for simultaneous lineups across all levels of suspect position.

A closer look at the difference between high- and low-similarity neighbor conditions showed that sequential low-similarity neighbor lineups increased correct and false identifications compared to sequential high-similarity neighbor lineups in early suspect positions, but discriminability did not improve. Thus, witnesses were more likely to identify the suspect when preceded by a low-similarity lineup member. Low-similarity simultaneous lineups yielded slightly more correct identifications and false identifications compared to high-similarity simultaneous lineups, but the differences between discriminability and identification rates by neighborhood conditions were small and not statistically significant.

Taken together, the results suggested that simultaneous lineups were more accurate than sequential lineups, especially when the suspect appeared in position 2 surrounded by high-similarity lineup members, providing partial support for Wixted and Mickes' (2014) theory that comparisons between lineup members underlie the simultaneous advantage. The present study did not replicate the advantage for high-similarity neighbors compared to low-similarity neighbors in simultaneous lineups, nor did it replicate the confidence-accuracy inversion, possible reasons for which are discussed below. However, when comparing simultaneous to sequential lineups, high-similarity neighbors improved accuracy in suspect position 2. Presumably, it is easier to compare the best matching lineup members to determine which features are irrelevant, and which are diagnostic of guilt in high-similarity simultaneous compared to sequential lineups. At a glance, witnesses can see that all of the best matching lineup members are at the beginning of the lineup in the simultaneous condition, whereas in the sequential lineup, witnesses have no way of knowing whether there will be better alternatives beyond the second lineup member. The simultaneous advantage is attenuated when low-similarity lineup members surround the suspect, because preceding the suspect with a low-similarity foil increases the correct identification rate in sequential lineups.

Methodological Considerations

Unlike in Experiment 2, the high-similarity neighbor condition did not improve accuracy compared to the low-similarity neighbor condition within simultaneous lineups. In fact, low-similarity neighbor lineups produced slightly more correct and false identifications compared to high-similarity neighbor lineups. What accounts for these differences between experiments?

As mentioned earlier, an important factor that differentiated Experiments 2 and 3 was the similarity of the high-similarity lineup members. In Experiment 2, the highest similarity and second-highest similarity lineup members were rated as having an average similarity of 3.47 and 3.22 to the guilty suspect, whereas the highest and second highest-similarity lineup members were rated as having an average similarity of 3.22 and 3.11 to the guilty suspect in Experiment 3. Decreasing the similarity of the high-similarity lineup members may have removed useful comparative information, as the second highest similarity lineup member from Experiment 3 may have shared fewer features with the guilty suspect compared to the highest-similarity lineup member from Experiment 2. Further, the confidence-accuracy inversion in Experiment 2 was not replicated in Experiment 3. If the perceived difficulty account of the confidence-accuracy inversion was correct, then it makes sense that confidence would not invert in the present study, as the lineup members were less similar to the suspect compared to Experiment 2, making the task seem easier.

Fitzgerald et al. (2015) found that moderately high similarity lineups improved accuracy relative to very high similarity lineups. One could speculate that the advantage of high-similarity neighbors in Experiment 2 was small because the lineup members were too similar, mirroring Fitzgerald's et al.'s very high similarity condition. In hindsight, however, Fitzgerald et al. used morph faces to create very high and moderately high conditions, thus, their very high-similarity foils may have been much more similar to their suspect compared to the high-similarity foils and suspect in the present studies. Rather, the higher-similarity foils used in Experiment 2 may have been closer (in terms of similarity) to Fitzgerald's moderately high similarity lineups than to their very high similarity lineups. Thus, when using real rather than morph faces, the foils must be highly similar to the suspect to improve performance. Future studies on the neighborhood effect should use morph faces to control the level of similarity between the foils and the suspect, and establish the level of similarity required to produce the high-similarity neighbor advantage.

The difference between the results of the simultaneous lineups in Experiments 2 and 3 may also have been due to subtle methodological differences. The simultaneous condition in Experiment 3 was not an exact replication of Experiment 2. An exact replication uses the same methods to replicate a study, whereas Experiment 3 made changes in the implementation of the study that were considered to be unrelated to theory. Changes in the implementation of studies have been shown to yield variation in effect sizes, reducing general replicability (McShane & Böckenholt, 2015). As described in the methods section, Experiment 3 used different confidence scales, filler tasks, and instructions compared to Experiment 2. Further, because Experiment 3 was computerized, participants had less interaction with research assistants compared to Experiment 2.

Theoretical Considerations

The results provided some support for the hypothesis that comparisons between lineup members underlie the simultaneous advantage (Wixted & Mickes, 2014). The simultaneous lineup allows witnesses to compare the lineup members all at once; the sequential lineup does not. When high-similarity lineup members surround the suspect in simultaneous lineups, witnesses can take advantage of the proximity of the best-matching group of lineup members to make comparisons between relevant lineup members, improving performance relative to the sequential condition. The high-similarity simultaneous advantage was specific to suspect position 2, as performance declined in general across levels of suspect position in simultaneous lineups.

A review of the literature on suspect position effects by Clark et al. (2013b) suggested that position effects are seldom reported in simultaneous lineups for several reasons: some studies control for suspect position, others find no impact of suspect position, or studies are underpowered, and breaking up identification rates by position yields small cell sizes, resulting in meaningless comparisons. Studies that do report position effects have shown that witnesses tend to make identifications of lineup members in the middle positions of the lineup (Mecklenburg, 2006; O'Connell & Synnott, 2009; Sporer, 1993). The present result is more consistent with Gronlund et al.'s (2009) findings, which yielded a simultaneous lineup advantage when the suspect was in position 2 (top middle position in a 2 × 3 array). Although the lineup format between the present study and Gronlund et al.'s study is very different, the effects are parallel. Specifically, Gronlund et al. found that correct identifications were higher in simultaneous lineups in suspect position 2 compared to sequential lineups. Placing the suspect in position 5 (bottom middle) in a simultaneous lineup resulted in a sequential advantage. When the suspect was shifted to position 5, both the correct identification rate decreased and the false identification rate increased in simultaneous lineups. Because overall accuracy decreased in simultaneous lineups, whereas the suspect identification rates for sequential lineups remained fairly constant across positions, sequential lineups maintained an advantage via a simultaneous position effect in Gronlund et al.'s study—similar to the findings in the present study, in which the correct identification rates generally declined in later suspect positions, and false identifications generally increased in later suspect positions. Gronlund et al. speculated that the position effect was the product of neighborhood effects.

Neighborhood effects cannot explain the present results, however, as the impact of neighborhood was attenuated in position 4 in Experiment 2, and across position in Experiment 3 within simultaneous lineups. In the present studies, the lineup was presented in a horizontal line, and witnesses may have processed the lineup from left-to-right. After examining the first few lineup alternatives, witnesses may have relied on less effortful decision strategies, avoiding the comparative processes that impacted performance in suspect position 2.

Neighborhood effects were quite different in sequential lineups. In the present study, having high-similarity neighbors early in sequential lineups reduces discriminability compared to low-similarity neighbors, consistent with Clark and Davey's (2005) finding that the correct identification rate was higher when the next-best foil appeared after the suspect. When the next-best foil appeared earlier in the lineup, participants tended to spend their choice on a foil; thus, the correct identification rate decreased. In contrast, placing lower similarity foils prior to the suspect increased the correct identification rate—similar to the present study.

In sequential lineups, the low-similarity neighbor correct identification rate advantage may relate to perceptual aftereffects, which are perceptual changes that occur after viewing, or adapting to a stimulus. For example, Leopold, O'Toole, Vetter, and Blanz (2001) showed participants a face for a few seconds (called an adapting face) and then tested them on a different face. Leopold et al. showed that adaptation to an anti-face (that had the opposite features of the test face) improved identification performance on the test faces. In other words, viewing the anti-face increased sensitivity to the study face. Adaptation seems to alter the balance of responsiveness of groups of neurons that process facial structure. In the low-similarity neighbor sequential condition, perhaps witnesses adapted to the low-similarity foil, enhancing the similarity of the guilty suspect's features to their memory of the perpetrator, thereby increasing the correct identification rate.

Limitations and Future Research

The failure to replicate the high-similarity neighbor advantage within simultaneous lineups highlights the necessity of further work on the impact of neighborhood. What level of similarity produces the high-similarity advantage? Is there a level of similarity that is too similar? Is the neighborhood effect real? In addition to examining the effect of neighborhood across stimulus sets, lineup formats, instructions, and viewing conditions, it is critical to vary neighborhood similarity in future studies. Use of morph faces may allow for a controlled examination of similarity and the neighborhood effect.

The present study showed that suspect position impacted identification performance differently in simultaneous and sequential lineups—and that the impact of suspect position depended critically on the similarity of the foils preceding the suspect in sequential lineups. This finding reconciles results showing improvements in performance when the suspect appears in later positions (Gronlund et al., 2009) versus results showing improvements in performance when the suspect appears in early positions (Clark & Davey, 2005). The effect of the position of the suspect depended on the similarity of the nearby foils, both temporally and spatially. Given that suspect position is often controlled for or unreported in eyewitness identification research (Clark et al., 2013b), future studies should further explore the impact of position and foil similarity, as findings have practical application for lineup construction by law enforcement.

Experiment 3 joined several other recent studies showing that sequential lineups are not more accurate than simultaneous lineups. The factors underlying the simultaneous advantage are still not well understood. The present study provided some evidence in support of the idea that comparisons among lineup members underlie the simultaneous advantage, though this conclusion was tempered by the lack of a high-similarity neighbor advantage within simultaneous lineups. Given that sequential lineups are now a fixture of lineup administration in several states, an understanding of the impact of lineup procedure, suspect and foil similarity, and suspect position on accuracy is critical.

Chapter 5: General Discussion

The general discussion comprises four sections. The first section describes the current research on decision-making in eyewitness identification. The second section summarizes the results of Experiments 1-3. The third and fourth sections discuss the theoretical and practical implications of the results, respectively. The fifth section describes the limitations of the present studies and discusses future research directions.

Research on Eyewitness Decision-Making

The absolute-relative distinction has dominated eyewitness identification decision-making research for over 30 years. Absolute judgments—making an identification if the lineup member's match to memory exceeds a criterion—were assumed to underlie the so-called sequential advantage. In contrast, simultaneous lineups were assumed to encourage use of relative judgments—identifying the best match to memory relative to the other lineup members—yielding more false identification errors and lower overall accuracy. Relative judgments were proposed to be operative in simultaneous lineups because simultaneous lineups allow for comparisons between lineup members.

The finding of a sequential advantage is debated, however, with numerous recent studies showing that sequential lineups produce similar (Gronlund et al. 2012) or worse performance compared to simultaneous lineups (Mickes, et al. 2012, Dobolyi & Dodson, 2013, Carlson & Carlson, 2014). One of the main differences between research showing sequential superiority and recent research showing sequential inferiority derives from the use of different measures of accuracy. Early studies relied on the problematic C/F ratio as a measure of accuracy, whereas recent studies use ROC analyses (Wixted & Mickes, 2012). Further, studies showing a sequential advantage tended to hold suspect position constant (Lindsay, Lea, & Fulford, 1991; Lindsay, Lea, Nosworthy, et al., 1991) often in later positions (Malpass & McQuiston, 1999; Beresford & Blades, 2006) or provided no information about suspect position (Cutler & Penrod, 1988), ignoring a variable that influences the accuracy of sequential lineups (Clark et al., 2013b).

Most research on absolute and relative judgments is tied to the simultaneous/sequential paradigm in eyewitness identification (Lindsay & Wells, 1985; Wells, 1993) and face recognition

(Weber & Brewer, 2004) research. If sequential lineups are less accurate compared to simultaneous lineups, do absolute judgments produce worse accuracy? The conflicting evidence regarding the sequential lineup highlights the importance of examining absolute and relative judgments using different paradigms.

Researchers have criticized the absolute-relative distinction for a lack of operational definitions and noted the need for computational models to clarify aspects of the absolute-relative distinction (Clark, 2008; Clark et al., 2011; Gronlund, Mickes, Wixted, & Clark, 2015). Clark, Erickson and Breneman (2011) modeled identification response probabilities for absolute and relative judgments using the WITNESS model. According to WITNESS, absolute judgments put all of the decision weight on the best match and none on the difference between the best match and the next best match, $1.0(\text{BEST}) + 0.0(\text{BEST - NEXT}) < C$. Relative judgments put all of the decision weight on the difference between the best match and the next best match and none on the best match, $0.0(\text{BEST}) + 1.0(\text{BEST} - NEXT) < C_{\text{diff}}$.

Even with formal definitions of absolute and relative judgments, relative judgments may still be conceptualized in different ways. Some definitions of relative judgments seem to promote a response bias view (identifying the best matching lineup member) whereas others frame relative judgments in terms of a comparative process. Wixted and Mickes (2014) recently advanced a signal detection-based theory that predicts the simultaneous advantage based on comparisons between lineup members. To what extent do comparisons between lineup members underlie relative judgments? Few previous studies have examined absolute-and relative judgments outside of the simultaneous-sequential debate, and fewer have examined comparisons between lineup members in their own right. The present studies were designed to examine the impact of comparative processes on eyewitness identification accuracy using novel paradigms.

Summary of Results from Experiments 1-3

Experiment 1 provided an empirical test of the WITNESS model predictions of an absolute advantage. The research questions were: (1) What is the impact of absolute and relative judgments, as defined by the WITNESS model? (2) How do foil selection strategies, decision strategies, and suspect position interact?

Contrary to the prediction that absolute judgments produce higher accuracy, results suggested that absolute judgments did not produce higher accuracy than relative judgments. In fact, accuracy was slightly higher for participants who were instructed to make relative judgments compared to participants who were instructed to make absolute judgments according to identification results, ROC and d' analyses for both same-foils and different-foils lineups. The advantage for relative judgments varied by suspect position, and was most pronounced when the suspect appeared in position 6. The variation of the impact of relative judgments by suspect position could be attributed to the suspect's neighborhood. Higher accuracy for relative judgments was found in positions in which neighborhood similarity was higher, and a small disadvantage for relative judgments was found in positions in which neighborhood similarity was lower.

The results suggest that relative judgments do not necessarily increase false identifications, contrary to Wells' (1984) prediction based on the absolute-relative distinction and Clark et al.'s (2011) predictions based on the WITNESS model. The results are consistent with Wixted & Mickes (2014) prediction that the comparisons between lineup members allowed by relative judgments should increase accuracy.

The primary questions addressed in Experiment 2 concerned the effect of comparisons between lineup members on eyewitness identification accuracy and confidence. Does the proximity of highly similar foils to the suspect facilitate comparisons between lineup members, thus improving accuracy? How do comparisons between high-similarity lineup members and the suspect affect confidence and accuracy?

Consistent with Tulving's (1981) findings, the results showed that surrounding the suspect with high-similarity lineup members produced a small accuracy improvement but a decrease in confidence compared to surrounding the suspect with low-similarity lineup members. The finding owed to higher numbers of low confidence correct identifications in the high-similarity condition. Witnesses may have found the high-similarity neighbor condition more difficult than the low-similarity neighbor condition, despite their accurate performance. The results of Experiment 2 are also consistent with Wixted and Mickes' (2014) account of the simultaneous accuracy advantage, as well as with previous research showing similar lineups produce better identification performance compared to dissimilar lineups (Clark, 2012; Fitzgerald et al., 2013). Comparing highly similar lineup members facilitated performance relative to comparing dissimilar lineup members.

Experiment 3 investigated the underpinnings of the simultaneous advantage, addressing three main questions: (1) Is the accuracy advantage for simultaneous lineups due to comparative processes? (2) Are comparisons between lineup members, and the simultaneous accuracy advantage, dependent on lineup composition and arrangement? (3) What is the impact of suspect position on identification accuracy in simultaneous and sequential lineups?

Experiment 3 showed that comparisons between best matching lineup members improved accuracy more in simultaneous compared to sequential lineups, especially in suspect position 2, although there was no high-similarity advantage overall. The simultaenous advantage diminished across suspect position, as sequential discriminability increased when the suspect appeared later in the lineup. The low-similarity neighbors condition increased accuracy in simultaneous over sequential lineups as well, but to an attenuated degree compared to the high-similarity neighbors condition. Low-similarity lineup members increased both the correct and the false identification rate in sequential lineups, consistent with previous research showing that the position and relative similarity of the foils impacts performance differentially in sequential lineups (Clark & Davey, 2005). Experiment 3 failed to replicate the high-similarity neighbors accuracy advantage and confidence-accuracy inversion of Experiment 2, potentially because the similarity of the high-similarity neighbors was lower in Experiment 3.

The results of Experiment 3 suggested that comparisons between lineup members may underlie the simultaneous advantage, although this was most pronounced when the suspect appeared in early positions. Placing the suspect in later lineup positions improved sequential performance relative to placing the suspect in earlier positions, particularly when a low-similarity lineup member preceded the suspect. In contrast, placing the suspect in later horizontal positions in simultaneous lineups may have reduced witness' reliance on effortful strategies, thus dissipating the beneficial impact of comparisons between best-matching lineup members.

Theoretical Implications

According to Wixted and Mickes (2014), the original formulation of the absolute-relative distinction is best characterized as a theory of response bias. When witnesses use relative judgments, they will have a tendency to identify the best matching lineup member, regardless of the magnitude of the match value (Wells, 1984). In Experiment 1, relative judgments were instantiated according to this original formulation, instructing witnesses to make an identification if the best match to memory was sufficiently better than anyone else in the lineup, a direct restatement of the decision rule from the WITNESS model. Contrary to the longstanding assumption that relative judgments increase identification errors, and therefore increase bias to make any identification, results from Experiment 1 showed that relative judgments as defined by the WITNESS model might actually improve discriminability, although only to a small degree. This result is inconsistent with a theory of response bias, because correct identifications were higher and false identifications were slightly lower for relative instructions. These results are consistent with Wixted and Mickes' theory that relative comparisons between lineup members enhance discriminabilty in simultaneous lineups.

Experiment 2 examined comparisons between lineup members in isolation of absolute and relative decision rules. Experiment 2 manipulated the ease with which witnesses could compare the most similar lineup members by surrounding the suspect with high or low-similarity lineup members. The (small) improvement in discriminability for high-similarity neighbors is also consistent with Wixted & Mickes' (2014) theory.

Experiment 3 compared simultaneous and sequential lineups for high- and low-similarity neighbor conditions. Simultaneous lineups yielded higher accuracy compared to sequential lineups for both neighborhood conditions, although the effect was more pronounced for high-similarity neighbor lineups. High-similarity neighbor simultaneous lineups were more accurate than high-similarity neighbor sequential lineups when the suspect appeared in position 2. This accuracy advantage did not maintain over suspect position, as the false identification rate increased and the correct identification decreased across suspect position in simultaneous lineups, and the false identification decreased somewhat and the correct identification rate increased across suspect position in sequential lineups.

The decrease in discriminability in simultaneous lineups across suspect position may relate to the left-to-right format. Witnesses may have made useful comparisons between the first few lineup members but relied less on this strategy as the lineup progressed. The results from Experiment 3 suggest that comparisons between lineup members may underlie the simultaneous advantage to an extent, although the utility of this strategy may depend on the position of the suspect coupled with the format of the lineup (horizontal versus a 2×3 array). Even more critically, the similarity of the high-similarity neighborhood was lower in Experiment 3 compared to Experiment 2, which may have decreased the utility of a comparative strategy. The high-similarity foils may have had fewer features in common with the perpetrator.

Performance on sequential lineups was affected by the similarity of the preceding lineup member. Witnesses made more correct and false identifications when the suspect was preceded by low-similarity lineup members in positions 2 and 3, compared to high-similarity lineup members. The present results are consistent with previous studies suggesting that witnesses do not use a pure absolute judgment strategy in sequential lineups (Clark & Davey, 2005; Gronlund et al., 2009; Carlson et al., 2008). The similarity of the preceding lineup member would not affect performance if participants relied only on the match to memory with no contribution of the other lineup members. The WITNESS model allows for a mixture of absolute and relative strategies by placing some weight on both absolute and relative processes (Clark, 2003; Clark et al., 2011), thus, the present results may be explained in the WITNESS model by increasing the relative weight and decreasing the absolute weight for sequential lineups across suspect position.

Taken together, the results from the present studies suggest that information from comparisons between lineup members can be helpful. Relative judgments improved accuracy to a small extent, and absolute judgments did not improve accuracy. In simultaneous lineups, comparisons between the highest similarity lineup members were helpful, whereas in sequential lineups, viewing a few faces before the suspect may have improved the witness' ability to detect the suspect—though preceding the suspect with a low-similarity foil increased the likelihood of identifying both the innocent and the guilty.

Practical Implications

In general, the results suggest that policy recommendations to adopt a lineup procedure that prevents or attenuates comparisons between lineup members (e.g. sequential lineups or show-up procedures) should not be followed. This conclusion derives from three main findings: (1) When instructed to use a relative judgment (and therefore use comparative information to make their decision), witnesses performed better than when instructed to use an absolute judgment. (2) When the arrangement of the foils facilitated comparisons between highly similar lineup members, performance improved relative to arrangements that made those comparisons more difficult. (3) Simultaneous lineups, which allow for comparisons between lineup members, yielded an advantage over sequential lineups, especially when high-similarity foils surrounded the suspect in position 2. In all cases, the condition that encouraged comparisons between lineup members produced the best identification performance.

The results also have implications regarding the position of the suspect. Relatively little is known about how police officers determine the position of the suspect. Wogalter, Malpass, and McQuiston (2004) found that police officers reported that they generally placed the suspect in middle positions. Clark et al. (2013b) analyzed position data from an Illinois field study comparing simultaneous and sequential lineups (Mecklenburg, 2006) and found that suspect placements tended to be earlier in sequential lineups and later in simultaneous lineups–which were positions that decreased accuracy in sequential and simultaneous lineups (respectively) in Experiment 3. One should not conclude, based on the present results, that law enforcement should use late positions for sequential lineups and early positions for simultaneous lineups. More research should be conducted on the relationship between suspect position, foil similarity, and lineup presentation method to determine the impact of these three variables on accuracy.

Limitations and Future Directions

The present studies provide interesting insights regarding comparative processes underlying eyewitness identifications; although conclusions about the benefits of comparisons between lineup members are tempered by several limitations. In Experiment 1, presenting absolute judgments after making the best-match determination may have attenuated the effect of the instruction manipulation. Witnesses may have used comparative or relative information to make the best-match determination. It is possible that the relative advantage would have been stronger if witnesses were given the instruction prior to the onset of the lineup—or—the absolute instruction may have improved accuracy without making the best-match determination. Future studies will vary the timing of the instruction manipulation. Further, an extension of the present study to examine the effect of absolute and relative decision rules across simultaneous and sequential lineups and neighborhood conditions may be informative. If relative judgments improve performance, discriminability should be highest for simultaneous high-similarity lineups when witnesses are instructed to use a relative decision rule. Such a manipulation would also clarify the degree to which people use relative information in sequential lineups.

A second limitation of Experiment 1 involves the positions of the foils, which were fixed with respect to the suspect. This positioning strategy may have created configurations in the lineup that resulted in neighborhood effects, (although Experiments 2 and 3 show small and inconsistent neighborhood effects). Indeed, the suspect was identified more frequently in the relative condition in suspect position 6, which was the position that yielded the highest similarity between nearby foils and the suspect. The results suggest that studies of eyewitness identification should randomize the position of the suspect to control for potentially confounding lineup configurations.

Experiment 1 sampled participants online, which might have introduced a substantial amount of noise into the data. Participants could use any computer or device they preferred and could complete the study in any environment. Occasionally, the program “broke" causing the participant to miss some or all of the video. Substantial effort was made to find and drop all cases in which the participant did not witness the staged crime, but even a participant who provided a description of the crime like “a robbery," may not have seen the crime. Future studies will include a multiple-choice test at the end of the experiment that queries the participant on several details of the staged crime. Those who score poorly on the test will be dropped from analyses.

Experiment 3 did not replicate the high-similarity neighborhood advantage found in Experiment 2. Was the difference in results due to the decrease in the similarity of one of the lineup members? Was the effect of neighborhood in Experiment 2 due entirely to the presence of the highest-similarity foil? Or, was the neighborhood effect in Experiment 2 a fluke? A careful comparison of neighborhood effects with different levels of similarity of the neighboring foils is warranted. If the neighborhood effect was real, then the generality of the neighborhood effect needs to be established across variations in stimulus materials. Experiments 2 and 3 constructed the lineups as a horizontal row, but most lineups made by law enforcement are presented via a 2 × 3 array. It is critical to determine how neighborhood impacts performance in the latter type of lineup. Based on the results of Experiment 1, which showed that the effect of the instructions manipulation differed depending on the position of the suspect, there is reason to believe that configurations impact performance in the 2 × 3 array.

The current studies were the first to directly examine comparative processes independent of the simultaneous/sequential procedures. Other types of manipulations of comparative processes should be employed in future studies to corroborate the current findings. One way to examine the effect of comparisons between lineup members would involve going back to basics to examine hypotheses within a recognition memory context.

For example, a study adapted from Tulving (1981) would be informative. Tulving found higher accuracy, but lower confidence for recognition memory test trials that presented similar target-distractor pairs (A/A') compared to dissimilar target-distractor pairs (A/B'). Tulving's explanation of these results was that similar target-distractor pairs improve accuracy by allowing participants to determine the relevant, diagnostic features. Clark (1997) modeled Tulving's confidence-accuracy inversion by assuming that the match values for targets and distractors were correlated in the A/A' test trials. The correlation reduces the variance of the difference distribution for A/A' trials (because both the target and distractor match the memory of the target) but not for A/B' trials. The lower variance in the denominator of the signal-to-noise ratio yields higher accuracy. Clark's account does not require assumptions about comparative processes.

An informative study that would examine comparative processes would use a recognition memory task like Tulving's (1981) study. After studying a series of complex photographs, participants would be tested on similar versus dissimilar distractor test trials. Critically, the target-distractor pairs will be presented simultaneously or sequentially. If people make comparisons between stimuli to identify and focus on unique diagnostic features and ignore non-diagnostic shared features, then performance will be highest on simultaneous A/A' trials. The benefit of this comparison process should disappear when stimuli are presented sequentially, resulting in lower accuracy for A/A' sequential trials than for simultaneous trials. Clark's (1997) correlated familiarity model could not account for this result because the model assumes that familiarity does not change depending on item presentation.

The present studies raise interesting questions about the relationship between relative judgments, comparative processes, and accuracy. The degree to which relative judgments imply comparisons between lineup members is still unclear. This work echoes calls from other researchers regarding the need for formal theories of eyewitness identification (Gronlund et al., 2015; Clark, 2008; Clark et al., 2011; Fife et al., 2014) that clearly specify the relationship between relative judgments and comparisons between lineup members. Despite the limitations, the current work yielded a striking finding that contradicted the dominant theory in eyewitness identification: relative judgments, and comparisons between lineup members, improve eyewitness identification decisions.

References

Beresford, J., & Blades, M. (2006). Children's identification of faces from lineups: The effects of lineup presentation and instructions on accuracy. *Journal of Applied Psychology*, *91*, 1102-1113.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*, 26-48.

Carlson, C. A., & Carlson, M. A. (2014). An evaluation of lineup presentation, weapon presence, and a distinctive feature using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *3*, 45-53.

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

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629-654.

Clark, S. E. (2008). The importance (necessity) of computational modeling for eyewitness identification research. *Applied Cognitive Psychology*, *22*, 803-813.

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science*, *7*, 238-259.

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

Clark, S. E., Erickson, M. A., & Breneman, J. S. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, *35*, 364-380.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37-60.

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior*, *32*, 187-218.

Clark, S. E., Moreland, M. B., & Gronlund, S. D. (2013a). Evolution of the empirical and theoretical foundations of eyewitness identification reform. *Psychonomic Bulletin and Review*.

Clark, S. E., Rush, R., & Moreland, M. B. (2013b). Constructing the lineup: Law, reform, theory, and data. In B. L. Cutler (Ed.), *Reform of eyewitness identification procedures.* Washington DC: American Psychological Association.

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

Cutler, B. L., & Penrod, S. D. (1988). Forensically relevant moderators of the relationship between eyewitness identification accuracy and confidence. *Journal of Applied Psychology*, *74*, 650-653.

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied*, *19*, 345-357.

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818-835.

Ebbesen, E. B., & Flowe, H. (2002). Simultaneous v. sequential lineups: What do we really know? *Unpublished Manuscript*.

Fife, D., Perry, C., & Gronlund, S. D. (2014). Revisiting absolute and relative judgments in the WITNESS model. *Psychonomic Bulletin & Review*, *21*, 479-487.

Fitzgerald, R. J., Oriet, C., & Price, H. L. (2015). Suspect-filler similarity in eyewitness lineups: A literature review and a novel manipulation. *Law and Human Behavior*, *39*, 62-74.

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*, 151-164.

Flowe, H. D., & Ebbesen, E. B. (2007). The effect of lineup member similarity on recognition accuracy in simultaneous and sequential lineups. *Law and Human Behavior*, *31*, 33-52.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton-Century-Crofts: New York, NY: Appleton-Century-Crofts.

Gonzalez, R., Davis, J., & Ellsworth, P. C. (1995). Who should stand next to the suspect? Problems in the assessment of lineup fairness. *Journal of Applied Psychology*, *80*, 525-531.

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior*, *34*, 445-459.

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika*, *32*, 25-33.

Gronlund, S. D., Andersen, S. M., & Perry, C. (2013). Lineup presentation. In B. L. Cutler (Ed.), *Reform of eyewitness identification procedures.* Washington DC: American Psychological Association.

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

Gronlund, S. D., Carlson, C. A., Neuschatz, J. S., Goodsell, C. A., Wetmore, S. A., Wooten, A., et al. (2013). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *1*, 221-228.

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. *Manuscript in preparation*.

Gross, S. R., & Shaffer, M. (2012). *Exonerations in the United States, 1989-2012* (Tech. Rep.). University of Michigan Law School.

Hintzman, D. L. (1993). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), (Relating theory and data: Essays on human memory in honor of Bennet B. Murdock ed., p. 39-56). Hillsdale, NJ: Erlbaum.

John, O. P., Donahue, E. M., & Kentle, R. (1994). *“The big five" inventory–version 4a and 54* (Tech. Rep.). University of California, Berkeley, Institute of Personality and Social Psychology: University of California, Berkeley, Institute of Personality and Social Psychology.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *22*, 1304-1316.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the “general acceptance" of eyewitness testimony research. *American Psychologist*, *56*, 405-416.

Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology*, *15*, 659-671.

Leopold, D. A., O’Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*, 89-94.

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

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., et al. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, *76*, 796-802.

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 and Human Behavior*, *21*, 391-404.

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

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A Users’s Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Malpass, D. C. . R. S., & McQuiston, D. (1999). Parallelism in eyewitness and mock witness identifications. *Applied Cognitive Psychology*, *13*.

Manson v. Brathwaite. (1977). 432 U.S. 98, 97 Sup. Ct. 2243.

Mansour, J. K., Lindsay, R. C. L., Brewer, N., & Munhall, K. G. (2009). Characterizing visual behavior in a lineup task. *Applied Cognitive Psychology*, *23*, 1012-1026.

McShane, B. B., & Böckenholt, U. (2015). You cannot step into the same river twice: When power analyses are optimistic. *Perspectives on Psychological Science*, *9*, 612-640.

Mecklenburg, S. H. (2006). Report to the legislature of the state of Illinois: The Illinois pilot program on sequential double-blind identification procedures. *Available at http://www.chicagopolice.org/IL%20Pilot%20on%20Eyewitness%20ID.pdf*.

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: Comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*, 361-376.

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute d', not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition*, *3*, 58-62.

Murdock, B. (1974). *Human memory: Theory and data.* Oxford England: Lawrence Erlbaum.

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

Neil v. Biggers. (1972). 409 U.S. 1988 93 Sup. Ct. 375.

O'Connell, M., & Synnott, J. (2009). Position of influence: Variation in offender identification rates by location in a lineup. *Journal of Investigative Psychology and Offender Profiling*, *6*, 139-149.

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247-255.

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, *19*, 55-71.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.

Ross, D. F., Benton, T. R., McDonnell, S., Metzger, R., & Silver, C. (2007). When accurate and inaccurate eyewitnesses look the same: A limitation of the 'pop-out' effect and the 10- to 12-second rule. *Applied Cognitive Psychology*, *21*, 677-690.

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

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychonomic Bulletin*, *118*, 315-327.

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

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *28*, 643-662.

Thompson-Cannino, J., Cotton, R., & Torneo, E. (2009). *Picking cotton: Our memoir of injustice and redemption*. New York, NY: St. Martin's Press.

Tulving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior*, *20*, 479-469.

U. S. Department of Justice, Office of Justice Programs, National Institute of Justice. (1999). *Eyewitness evidence: A guide for law enforcement.* Retrieved from `https://www.ncjrs.gov/pdffiles1/nij/178240`

U.S. v. Telfaire. (1972). 469 F.2d 522 (D.C. Cir.).

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology–Learning, Memory, and Cognition*, *26*, 582-600.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied*, *10*, 156-172.

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

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

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

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45-75.

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2014). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*.

Wickens, T. D. (2001). *Elementary Signal Detection Theory*. Oxford: Oxford University Press.

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*, 262-276.

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (In Press). Initial eyewitness confidence reliably predicts eyewitness identification accuracy.

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime, & Law*, *10*, 69-82.

Woocher, F. D. (1977). Did your eyes deceive you - expert psychological testimony on unreliability of eyewitness identification. *Stanford Law Review*, *29*, 969-1030.