

Probative Value of Absolute and Relative Judgments in Eyewitness Identification

Steven E. Clark · Michael A. Erickson ·
Jesse Breneman

Published online: 16 October 2010

© American Psychology-Law Society/Division 41 of the American Psychological Association 2010

Abstract It is well-accepted that eyewitness identification decisions based on relative judgments are less accurate than identification decisions based on absolute judgments. However, the theoretical foundation for this view has not been established. In this study relative and absolute judgments were compared through simulations of the WITNESS model (Clark, *Appl Cogn Psychol* 17:629–654, 2003) to address the question: Do suspect identifications based on absolute judgments have higher probative value than suspect identifications based on relative judgments? Simulations of the WITNESS model showed a consistent advantage for absolute judgments over relative judgments for suspect-matched lineups. However, simulations of same-foils lineups showed a complex interaction based on the accuracy of memory and the similarity relationships among lineup members.

Keywords Eyewitness identification · Memory · Decision-making

It is common in criminal investigations to present witnesses with lineups for the purpose of identifying, or excluding, a person who is suspected by the police. The lineup typically includes six or more individuals (or their photographs), one of whom is suspected by police; the others are foils or fillers who are known to be innocent. The witness may identify someone from the lineup, but may also decline to identify anyone.

It is well-known, of course, that witnesses sometimes make mistakes, and that these mistakes sometimes result in convictions of the innocent (Brandon & Davies, 1973; Gross, Jacoby, Matheson, Montgomery, & Patil, 2005; Huff, Rattner, & Sagarin, 1996; Scheck, Neufeld, & Dwyer, 2000; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). Such errors have motivated a large research literature focused on understanding the factors and underlying mechanisms that lead to eyewitness identification errors.

A fundamental question, and the focus of the present research, concerns the decision processes that witnesses employ in making eyewitness identification decisions. Wells (1984) placed a cornerstone in this area of research by distinguishing between two different decision strategies, one based on *absolute* judgments and the other based on *relative* judgments. In addition, Wells suggested that misidentifications of the innocent are due largely to relative judgments rather than absolute judgments. This distinction between absolute and relative judgments is the focus of the present research.

According to Wells' proposal, a witness using an absolute judgment makes an identification of a lineup member if the match between that lineup member and the witness's memory of the perpetrator is sufficiently high, above some criterion, whereas a witness using a relative judgment strategy makes an identification if the match of that lineup member is a relatively better match than any other lineup member. It is clear from this description that a witness utilizing a relative judgment strategy might identify a lineup member not because that person was a particularly close match to his or her memory of the perpetrator, but because that person was a better match than anyone else in the lineup.

Wells (1984, 1993) has suggested that absolute and relative decision rules produce equivalent patterns of results if

S. E. Clark (✉) · M. A. Erickson · J. Breneman
University of California, Riverside, Riverside, CA 92521, USA
e-mail: clark@ucr.edu

the perpetrator is in the lineup. However, in those cases in which the perpetrator is not in the lineup, Wells has described the relative judgment strategy as “fallacious” (Wells, 1984, p. 89), “problematic” (Wells, 1993, p. 553), “dangerous,” and “dysfunctional” (Wells, 1993, p. 560).

According to this view, the perpetrator’s presence or absence is a key consideration in evaluating absolute versus relative decision strategies. In the experimental literature, the perpetrator-present lineup condition simulates the real-world case in which the person the police suspect of having committed the crime did commit the crime, whereas the perpetrator-absent lineup condition simulates the real-world case in which the person suspected by the police is innocent of the crime. Suspect identifications are of central importance, and we will use the term *correct identification* to refer to an identification of the guilty suspect from a perpetrator-present lineup, and the term *false identification* to refer specifically to the identification of the innocent suspect from a perpetrator-absent lineup. According to Wells (1984), correct identification rates are unaffected by witnesses’ use of relative judgments, but false identifications—the specific error that can lead to a wrongful conviction—are increased by witnesses’ use of relative judgments. This view, that relative judgments are problematic only in perpetrator-absent lineups, but not perpetrator-present lineups, is widely held and oft-repeated (Dysart & Lindsay, 2007; Kneller, Memon, & Stevenage, 2001; Lindsay, 1999; Lindsay & Bellinger, 1999; Lindsay, Lea, & Fulford, 1991; Pozzulo, Crescini, & Lemieux, 2008).

This combination of non-effects for correct identification rates and increases for false identification rates leads to the conclusion that identification decisions based on absolute judgments are more accurate than identification decisions based on relative judgments. In a legal context, “accuracy” is sometimes described in terms of the probative value of the eyewitness identification evidence. A suspect identification has probative value to the degree that the likelihood of a correct identification is high and the likelihood of a false identification is low. Probative value is a legal term without a precise mathematical definition (Kaye, 1986). However, estimates of probative value are often calculated as the conditional probability of guilt given that the suspect was identified, i.e., $PV_{CP} = \text{correct} / (\text{correct} + \text{false})$, or by the ratio of correct to false identifications, i.e., $PV_{RATIO} = \text{correct} / \text{false}$. We use both measures, and in addition, we calculate d' (Tanner & Swets, 1954) from correct and false identification rates.

However probative value is measured, by a conditional probability, a ratio, or by d' , if absolute judgments are indeed better than relative judgments, then the probative value of a suspect identification should be higher when witnesses’ decisions arise from absolute, rather than relative, judgments. One might expect the next question to ask

whether this prediction is confirmed or disconfirmed by empirical data. However, we begin with a different, more fundamental, question:

What is the theoretical basis of these predictions? *Why* should the shift from absolute to relative judgments only affect false identification rates with little or no effect on correct identification rates? What is the basis of the view that absolute judgments are better than relative judgments?

It is important to note that the predictions described above, regarding correct and false identification rates, and the probative value of a suspect identification, are not derived from any theory of memory or decision-making. Rather, the absolute-relative distinction was motivated and shaped by a recurring pattern of results. Wells (1984) noted experimental results by Malpass and Devine (1981) and by Lindsay and Wells (1980) which did show increases in false identifications, with little or no increase in correct identifications, when witnesses were given biased instructions prior to seeing the lineup, and when the lineup contained foils that were not similar to the suspect. The absolute-relative distinction also motivated the development of the sequential lineup. In 1985, Lindsay and Wells published the first empirical comparison between sequential and simultaneous lineups, which showed the same asymmetric pattern: Simultaneous lineups produced a much higher rate of false identifications, with only a small, statistically insignificant increase in correct identifications.

In the mid-1980s, this pattern of results appeared to be a regularity of eyewitness identification. However, more recent and expansive reviews of the literature, by Clark, Howell, and Davey (2008) and by Clark and Godfrey (2009), suggest that this pattern is not the rule, but rather the exception. The pattern that is most often shown is one in which correct and false identification rates vary together. For example, averaged across studies that compared biased and unbiased instructions, correct identification rates increased from .495 for unbiased instructions to .559 for biased instructions, and false identification rates increased from .071 for unbiased instructions to .096 for biased instructions. These increases in correct and false identification rates combined to produce only a very small decrease in the probative value of a suspect identification, from .869 for unbiased instructions to .851 for biased instructions. Similarly, 20 of 29 simultaneous-sequential lineup comparisons reviewed by Clark and Godfrey showed a pattern of lower correct identification rates and lower false identification rates for sequential lineups relative to simultaneous lineups. Collapsed over all 29 studies, the probative value of a suspect identification showed a modest increase from .775 to .812. The bottom line is that the pattern described by Wells (1984) is obtained sometimes; however, the most common pattern of results shows changes in both correct and false identification rates, often

with little or no change in the probative value of a suspect identification.

What are the implications of these two patterns of results? Empirical results that show an asymmetrical increase in false identifications and an overall increase in probative value are often offered as evidence in support of the absolute-relative distinction and a disconfirmation of criterion-shift models (Stebly, Dysart, Fulero, & Lindsay, 2001; Wells, 1984). Conversely, empirical results that show proportional increases in correct and false identification rates, with no increase in the probative value of a suspect identification, have been offered as a challenge to the absolute-relative distinction and as a confirmation of a criterion-shift model (MacLin, Meissner, & Zimmerman, 2005; Meissner, Tredoux, Parker, & MacLin, 2005).

If the asymmetric pattern of results is viewed as a confirmation of the distinction between absolute and relative judgments and a disconfirmation of criterion-shift explanations, should the reverse also hold? In other words, do results that show proportional increases in correct and false identification rates, with no change in the probative value of a suspect identification, constitute a disconfirmation of the absolute-relative judgment distinction and a confirmation of the criterion-shift model?

In contrast to what seems like a straightforward empirical test of competing theories, the relationship between theory and data is not straightforward. The problem in assessing the implications of these two patterns of results is that the predictions of the two theoretical frameworks have not been derived or demonstrated to arise from any theory of memory or decision-making. The problem is not in the clarity of the predictions, as the predictions have been clearly, and repeatedly, stated. Rather, the problem is in the theoretical *foundation* of the predictions. What is the basis for the prediction that absolute judgments are better—that they lead to more accurate responding and identification evidence with higher probative value—than relative judgments? Does the absolute judgment advantage arise from any theory of memory or decision-making? If so, how general is it? What produces it?

We address these questions in this article by generating response probabilities according to absolute or relative decision rules, implemented within the framework of a single computational model. Thus, the decision rules need not be inferred from the responses of human witnesses. Instead the model generates the responses, and thus we can directly compare responses that are known to be generated by absolute or relative decision rules. This technique of comparing models directly—rather than inferring psychological processes from human data—is an increasingly common and useful approach in cognitive and social psychology (Gigerenzer & Goldstein, 1996; Hastie & Stasser, 2000; Raaijmakers & Shiffrin, 1981; Wixted, 2007). With

few exceptions (i.e., Penrod & Hastie, 1980), formal models (including computer simulation models) have been utilized infrequently in psychology and law (see Ogloff, 2000; Small, 1993), although that may be changing (see Clare & Lewandowsky, 2004; Clark, 2003; Goodsell, Gronlund, & Carlson, 2010; Gronlund, 2005; Swets, Dawes, & Monahan, 2000).

We next describe the implementation of absolute and relative decision rules within the framework of the WITNESS model. We then generate predictions of the model, across different lineup conditions, and across a wide range of the model's parameters, not only to establish the generality of the predictions, but more importantly to understand the behavior of the model and the basis of its predictions.

The WITNESS Model

WITNESS is a computational memory and decision model of eyewitness identification that generates response probabilities for suspect, foil, and nonidentification responses. The model represents information as vectors of features. Thus, the perpetrator is denoted as a vector \mathbf{P} , the witness's memory of the perpetrator as a vector \mathbf{M} , and the lineup members as vectors \mathbf{L}_1 , \mathbf{L}_2 , \mathbf{L}_3 , and so forth.

The specific features of the perpetrator need not be specified. Instead, each cell of the perpetrator vector is sampled from a rectangular, zero-centered, distribution. The specifics here are not critical. The elements of the vector could also be sampled from a discrete distribution (Hintzman, 1988), a normal distribution (Murdock, 1982), or a geometric distribution (Shiffrin & Steyvers, 1997). Indeed, these assumptions do vary across models that are very similar in their critical assumptions that (a) similar stimuli are represented by similar vectors, and (b) similarity is defined in terms of feature overlap (see Clark & Gronlund, 1996).

The memory trace of the perpetrator is assumed to be incomplete and to contain errors. Specifically, in the model, feature j of the perpetrator \mathbf{P} will, at the time of the lineup, be represented accurately in memory with probability a , and will be represented inaccurately (sampled randomly from a rectangular distribution) with probability $1 - a$. The memory accuracy parameter, a , varies as a function of the witness's limited opportunity to observe as well as the loss of information over time.

Similarity Relationships

The performance of the model is determined in large part by the similarity relationships in the lineup, specifically how similar the innocent suspect and the foils are to the

perpetrator. Similarity in the WITNESS model is given by the overlap between two vectors. All similarity relationships in the WITNESS model are represented in this way, so we will describe similarity in general terms first, and later show how it is applied in the context of a lineup. The similarity between any two vectors \mathbf{X} and \mathbf{Y} is given by a parameter $s(\mathbf{X}, \mathbf{Y})$. With probability $s(\mathbf{X}, \mathbf{Y})$, element j of vector \mathbf{X} will have the same value as element j of vector \mathbf{Y} , and with a probability of $1 - s(\mathbf{X}, \mathbf{Y})$, the value of element j of vector \mathbf{Y} is randomly determined.

The specific similarity relationships among the perpetrator, innocent suspect, and foils depend on how the lineup is constructed. An important distinction, which we explore in considerable detail, concerns the way in which lineups are created for laboratory experiments versus how they are created in real criminal investigations. In almost all eye-witness identification experiments, perpetrator-present and perpetrator-absent lineups contain the same foils. By contrast, police typically select foils based on the similarity to the suspect, who may be guilty or innocent (Wogalter, Malpass, & McQuiston, 2004). If the suspect is guilty the foils will be similar to that guilty suspect (i.e., the perpetrator); however, if the suspect is innocent, the foils will be similar to that innocent suspect. Hence, these suspect-matched lineups contain different sets of foils for perpetrator-present and perpetrator-absent lineups. This distinction between same-foils lineups and suspect-matched lineups is a critical factor in the comparison of absolute and relative judgment models. Consequently, the results for these two lineup conditions are presented separately, and the specific details for each lineup condition are presented within each separate section.

Comparison and Decision Processes

As the first step in the decision process, the WITNESS model computes the match between each lineup member and the witness's memory of the perpetrator. These match values are computed as the dot product between each lineup member vector \mathbf{L}_i and the memory vector \mathbf{M} . The dot product is then divided by the number of elements in the vectors. There are as many match values as there are lineup members, and all that is left is to generate a response based on some decision rule. The decision rule, of course, is the focus of this study. Various possible decision rules are discussed next.

According to the WITNESS model, an identification is made if the evidence in favor of an identification EV_{ID} exceeds a decision criterion c . Specifically, the best-matching lineup member will be identified if $EV_{ID} > c$, where $EV_{ID} = w_A(\text{BEST}) + w_R(\text{BEST} - \text{NEXT})$, here BEST is the value of the best match to memory, NEXT is the value of the next-best match to memory, c is the

decision criterion, and w_A and w_R are weights that must sum to 1.0.

Absolute and relative decision rules can be implemented in the model based on the values of w_A and w_R . If $w_A = 1.0$ (and thus $w_R = 0$), the decision rule is quite similar to the absolute decision rule originally described by Wells (1984); the identification is made if the match exceeds "some cut-off or threshold" (p. 95). We refer to this as the Best Above Criterion Model. Wells' description of absolute judgments did not specify the response that would be generated if more than one lineup member was above criterion. The Best Above Criterion Model allows an identification even if two or more lineup members are above criterion. Another reasonable implementation of an absolute judgment rule specifies that a positive identification is made if one, and only one, match is above criterion. We refer to this model as the One Above Criterion Model.

The model reduces to a pure relative decision rule if $w_R = 1$ (and $w_A = 0$). We refer to this as a *Relative Difference Model*. In this case the witness chooses "the lineup member who most resembles the witness's memory (of the perpetrator) relative to other lineup members" (Wells, 1984, p. 92). This model, which bases identification decisions on the NEXT-BEST difference, is a natural implementation of a relative judgment rule within the WITNESS model. However, it is not the only way to implement a relative judgment rule. Following the work of Sauer, Brewer, and Weber (2008), we considered another version of a relative difference model we call the BEST-REST model, where REST is the average of the other lineup members. (It is the equivalent of Sauer et al.'s $\text{MAX} - M_{\text{other}}$ decision rule.)

Simulations and Model Comparisons

Response probabilities were generated by computer simulation. Each simulation is the equivalent of one human subject, except that unlike human subjects, the model can specify the level of memory accuracy, the decision criterion, and the decision rule. Variation across simulations is produced by the probabilistic aspects of creating vectors and storing information in memory. Unlike human subjects, thousands of simulations can be run in just a few seconds, and hence each data point that we present is based on 6,000 simulations. Thus, the response probabilities produced by the model are very stable.

Correct and false identification rates will, of course, change as decision criteria are varied, for all models—the Best Above Criterion model, the One Above Criterion model, the BEST-NEXT model, and the BEST-REST

model. For example, the Best Above Criterion model will produce a very low overall identification rate if the criterion is set high, and the BEST-NEXT model will produce a very high overall identification rate if the criterion is set very low. For all models the decision criterion was varied from very high to very low, producing curves that share many of the properties of standard receiver operating characteristic (ROC) curves.

Suspect-Matched Lineups Versus Same-Foils Lineups

As noted earlier, two sets of simulations were conducted, one for same-foils lineups and another for suspect-matched lineups. Detailed discussions of same-foils and suspect-matched lineups are given elsewhere (Clark et al., 2008; Clark & Tunnicliff, 2001), and therefore, our discussion here is fairly brief.

Most laboratory experiments use a same-foils design, created by various means. The photograph of the perpetrator may be replaced by the photograph of a different person (who may assume the role of an innocent suspect), or the foils may be selected based on their match to a general description of the perpetrator. In either case, the foils in perpetrator-present and perpetrator-absent lineups are the same. By contrast, police officers typically construct lineups by selecting foils based on their similarity to their suspect (Wogalter et al., 2004). This selects different foils depending on whether the suspect is the perpetrator or an innocent person.

We conducted separate sets of simulations, for same-foils lineups and for suspect-matched lineups, for two reasons: (1) most experiments use a same-foils design, whereas most police officers select foils based on their similarity to the suspect. In order to fully understand the relationship between absolute and relative judgment models of eyewitness identification, and the role that relationship plays both in laboratory research and real-world criminal investigations, one must consider both lineup conditions—unless, of course, the distinction makes no difference, and (2) the distinction does make a very large difference.

Simulation Results

We turn now to the simulation results, presenting first the results from same-foils lineups and then the results from suspect-matched lineups. For both sets of simulations, we set the parameters in such a way as to produce a baseline level of “typical” performance, and then we varied the parameters in order to determine how variations from this baseline affected the output of the model.

Same-Foils Lineup Simulations

The simulations in this section consider the case in which the foils are the same in perpetrator-present and perpetrator-absent lineups. Because the foils are the same, their similarity can be described in terms of their relationship to the same target, i.e., the perpetrator. The similarity between a given foil and the perpetrator is determined by the parameter $s(F,P)$, and represents the proportion of features that are common between a given foil and the perpetrator. The innocent suspect, of course, may be more or less similar to the perpetrator than are the foils, and thus is represented by a separate parameter, denoted $s(I,P)$, i.e., the similarity of the innocent suspect to the perpetrator.

Baseline. Figure 1 shows the results of a simulation designed to produce typical eyewitness identification performance. As a measure of “typical” performance, we used the average correct and false identification rates from the recent meta-analysis by Clark et al. (2008). For all of the simulations, the vector length was set somewhat arbitrarily to 100 (as in previous applications of the model); $a = .3$, and $s(I,P) = s(F,P) = .6$. With these parameter values the model produced ROC curves that included the average correct and false identification rates from Clark et al. (2008), which are represented by the star symbol in the figure.

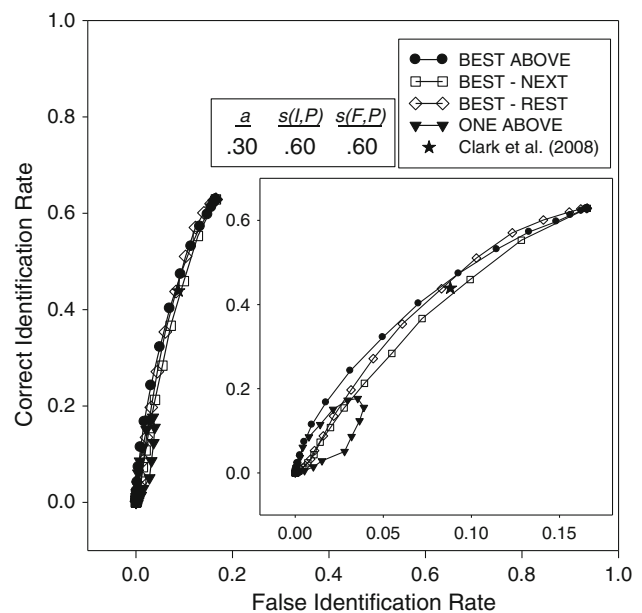


Fig. 1 Suspect ROC curves for same-foils lineups with baseline parameters, $a = .30$, $s(I,P) = .60$, and $s(F,P) = .60$. The vector length for this and all other simulations is 100. The star symbol represents typical results as the average correct (.443) and false identification (.085) rates from Clark et al. (2008)

The results are plotted in two ways. The main panel of the figure shows the ROC curves plotted in the traditional way, with the x and y -axes varying from 0 to 1. Plotted this way the curves for the Best Above Criterion, BEST-NEXT, and BEST-REST models are indistinguishable, implying equivalent performance across the three models. However, when the axes are stretched it is clear that the curves for those three models are close but not identical. There are advantages and disadvantages to each kind of ROC display. By keeping the axes constant, with a range of 0 to 1, the plot gives a true measure of the slope of the ROC curve. However, because the actual ROC curves do not stretch from 0 to 1, but rather are constrained in some cases to points between 0 and .167 (or less), the curves may be so compressed as to obscure real differences between the different decision models. Given that the focus of this article concerns *differences* in the decision rules, this is a necessary trade-off, without which large differences between the models can be hidden, leading to incorrect conclusions. Because the axes are stretched differently across figures, conclusions based on comparisons across figures should be made cautiously.

Before discussing the Best Above, BEST-NEXT, and BEST-REST models, it is useful to first discuss and dispense with the One Above Criterion model which produced an ROC curve sufficiently strange it can be eliminated from consideration. The reason for the odd performance is as follows: When the criterion is very high, no matches are above criterion, which leads to a nonidentification response. As the criterion is lowered, the suspect identification rate initially increases, but as the criterion continues downward, the suspect identification rate decreases. Why does the model produce this odd result? When the criterion is very high none of the lineup members are above criterion and thus a nonidentification response is generated. As the criterion is lowered, the likelihood that the suspect, *and only the suspect*, is above criterion increases, leading to the increase in the suspect identification rate. However, as the criterion is lowered further, the likelihood that two or more matches will be above criterion increases, violating the *one*-above decision rule, leading to an increase in non-identifications, and a decrease in suspect identifications. This odd performance arises from a general property of the model, such that the anomalous results shown in Fig. 1 are the norm, rather than the exception. Consequently, there is no need to discuss this model further.

With the One-Above Criterion model out of the way, we can now turn our attention to the performance of the Best Above Criterion, BEST-NEXT, and BEST-REST models. For all three models, as the decision criteria were decreased from very high to very low values, correct and false identification rates both increased, a result that is typically shown in ROC curves. However, the ROC curves differ

from those shown in yes–no signal detection tasks in one very clear way. Although the curves “grow” out of the lower left (0,0) origin, they do not project to the upper right (1,1) corner of the graph, but instead end rather abruptly. The reason for this is that correct and false identification rates are limited by the fact that the suspect is not always the best match to memory. The two curves project to the same point, a .643 correct identification rate and a .167 false identification rate. These are the probabilities that the perpetrator and innocent suspect are the best matches in their respective lineups. The .167 false identification rate, of course, is determined by the fact that the perpetrator-absent lineup is unbiased (because $s(I,P) = s(F,P)$).

Criterion Shifts and Measures of Probative Value. There is one aspect of the ROC curves that is not obviously apparent. For all three decision rules, increasing the decision criterion produces an increase in the probative value of a suspect identification. This point is illustrated in Table 1. The table shows the correct and false identification rates from three points on each ROC curve, for a low, medium, or high criterion, for the three remaining models. From these correct and false identification rates, three measures of probative value were calculated based on the conditional probability, ratio, and d' .

Table 1 Correct and false identification rates, probative value calculations, and d' at low, medium, and high criterion values, for best above criterion, BEST-NEXT, and BEST-REST models, for same-foils lineups. PV_{CP} is the probative value as a conditional probability (correct/correct + false); PV_{RATIO} is the correct/false ratio

	Low	Medium	High
Best above criterion			
Correct	.549	.424	.223
False	.123	.077	.028
PV_{CP}	.817	.847	.889
PV_{RATIO}	4.453	5.520	8.029
d'	1.281	1.236	1.153
Best-next			
Correct	.538	.383	.154
False	.123	.077	.028
PV_{CP}	.814	.833	.847
PV_{RATIO}	4.367	4.992	5.550
d'	1.254	1.131	.896
Best-rest			
Correct	.571	.414	.176
False	.123	.077	.028
PV_{CP}	.822	.843	.864
PV_{RATIO}	4.631	5.385	6.316
d'	1.337	1.209	.984

For all three models, the probative value of a suspect identification increases slightly when calculated as a conditional probability and it increases considerably when calculated as the ratio. Note for example that the probative value ratios from the Best Above Criterion model increase from 4.45 to 5.52 to 8.03 from the low to medium to high criterion values. This increase in probative value is surprising, as it has typically been assumed that criterion shifts have no effect on probative value. Indeed, they do, particularly when probative value is calculated as a ratio. By contrast, the variation in d' with criterion placement was much smaller. Changes in d' were the smallest for the Best Above Criterion model, and somewhat larger for the BEST-NEXT and BEST-REST models. When the underlying distributions are normal with equal variance, d' will not vary with the decision criterion. However, as will be shown, the underlying distributions, particularly for the BEST-NEXT and BEST-REST models, are not normal.

We turn now to the comparisons between the Best Above Criterion, BEST-NEXT, and BEST-REST models. Although the ROC curves for the three models appear very similar, they are not identical, a point that becomes clear when the axes are stretched as they are in the inset to Fig. 1. The inset shows that when the decision criteria are low there is a slight advantage for the BEST-REST model. To make comparisons across models, we can observe how the correct identification rate varies across models for a single, constant false identification rate. The correct identification rates, with the false identification rate at a constant value, are shown in Table 1. The table shows small, but reliable, differences across models. For example, for a constant false identification rate of .123, the correct identification rate for the BEST-REST model is roughly .02 higher than the Best Above Criterion model and .03 higher than the BEST-NEXT model. When the criteria are high, producing a false identification rate of .028, the pattern changes, with a correct identification rate for the Best Above Criterion model that is .07 higher than the BEST-NEXT model and .04 higher than the BEST-REST model.

The advantage for the Best Above Criterion model appears very small, almost imperceptible when the axes run from 0 to 1, but is apparent when the resolution is increased by adjusting the axes. Two questions are addressed next. First, how generalizable are the results shown in Fig. 1? The answer is that the results are not general at all, but vary as a function of memory accuracy and similarity relationships within the lineup. The second question is, What underlies the variability in the relative performance of the models? Each of these questions is addressed in turn.

Variation in Performance of Absolute and Relative Judgment Decision Rules

An extensive exploration of the model's parameter space showed an interaction of three factors: (1) the accuracy of memory, a ; (2) the similarity between the innocent suspect and the perpetrator, $s(I,P)$; and (3) the similarity between the foils and the perpetrator, $s(F,P)$.

The interaction is illustrated in Figs. 2 and 3. Figure 2 shows the ROC curves when memory was less accurate ($a = .2$), and Fig. 3 shows the ROC curves when memory was more accurate, $a = .6$. In both sets of simulations, the similarity parameters were varied with low and high values for $s(I,P)$ and low and high values for $s(F,P)$. In Fig. 2, $s(I,P)$ and $s(F,P)$ were set at .2 and .8, whereas in Fig. 3, $s(I,P)$ and $s(F,P)$ were set at .4 and .8. (The higher value of $s(I,P)$ was used because the lower value of .2 produced a nearly errorless ROC curve. We also conducted simulations with $a = .3$, and $s(I,P) = s(F,P)$ at .4 and .8, producing the same pattern of results shown in Fig. 2.) The 2×2 combination of similarity parameters is shown in panels A–D of Figs. 2 and 3. The results are summarized as follows.

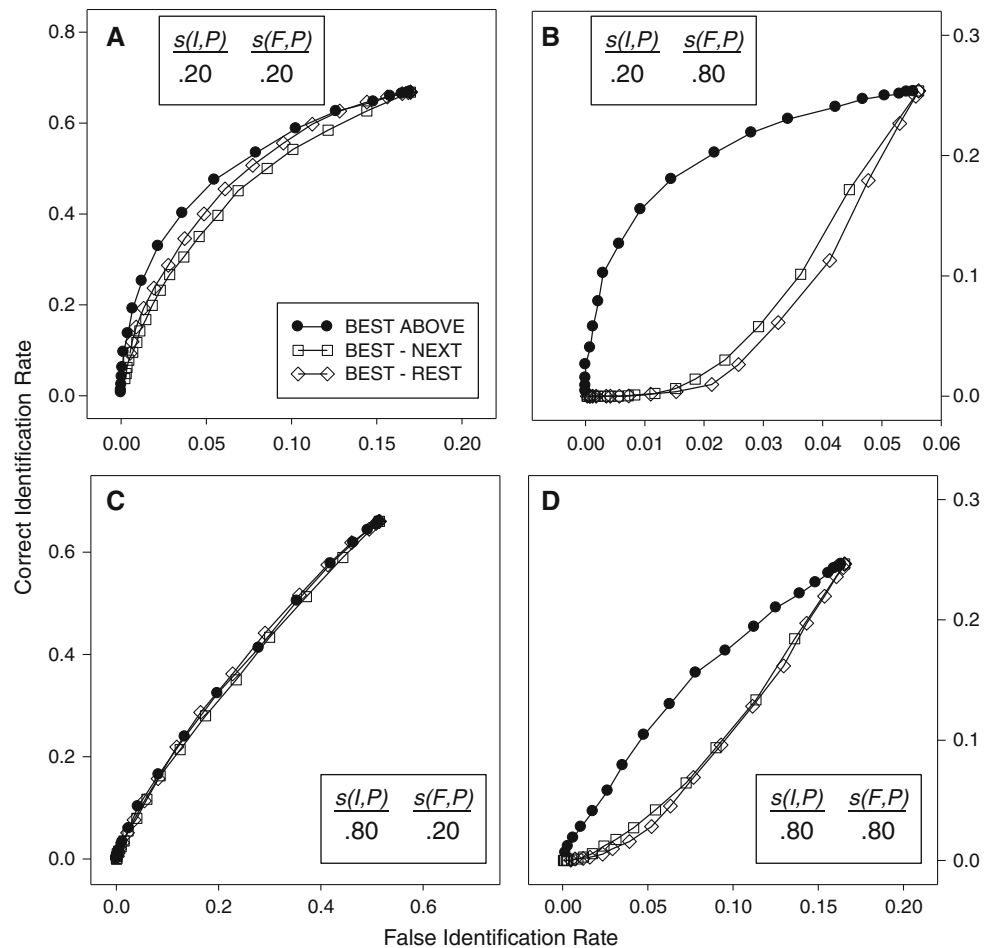
In general the Best Above Criterion model did better than the two relative difference models when memory was less accurate ($a < .4$). However, when memory was more accurate ($a > .5$) the BEST-REST model showed the highest level of performance, followed by the BEST-NEXT model, with the Best Above Criterion model showing the lowest performance. It is clear, however, that the accuracy of memory does not produce a main effect, but rather is a component of a complex interaction. The relative performance of the models depends not only on the value of a , but also on the values of the similarity parameters. The advantage for the Best Above Criterion model increases when the value of $s(F,P)$ is high (and the value of a is low). This is most apparent in panels b and d of Fig. 3.

The simulation results shown in Figs. 1–3 contradict Goodsell et al.'s (2010) conclusion that the different decision rules do not differentially affect discrimination between guilty and innocent suspects. Their conclusion was consistent with that of Breneman and Clark (2008). The differences in conclusions likely arise from our extended exploration of the parameter space as well as our stretching of the axes in plotting ROC curves.

Underlying Mechanisms

The patterns of results are, in some ways, quite sensible. The Best Above Criterion model does better than the two relative judgment models when the foils are very similar to the perpetrator because high values of $s(F,P)$ reduce the

Fig. 2 Suspect ROC curves for same-foils lineups with lower memory accuracy ($a = .2$) and 2×2 variation of low and high values for the similarity of the innocent suspect to the perpetrator, $s(I,P)$, and the similarity of the foils to the perpetrator, $s(F,P)$



BEST-NEXT and BEST-REST differences, which disproportionately harms performance of the relative judgment models. Conversely, the relative judgment models do better than the absolute judgment model when the innocent suspect is highly similar to the perpetrator because high values of $s(I,P)$ affect the *absolute* value of the match of the innocent suspect.

Of course, this only explains the similarity portion of the memory accuracy–similarity interaction shown in Figs. 2 and 3. Two other aspects of the results still require explanation. Why are the differences between the models often very small? And, how does the accuracy of memory affect the performance of the models? Answers to these questions require additional details regarding the behavior of the models.

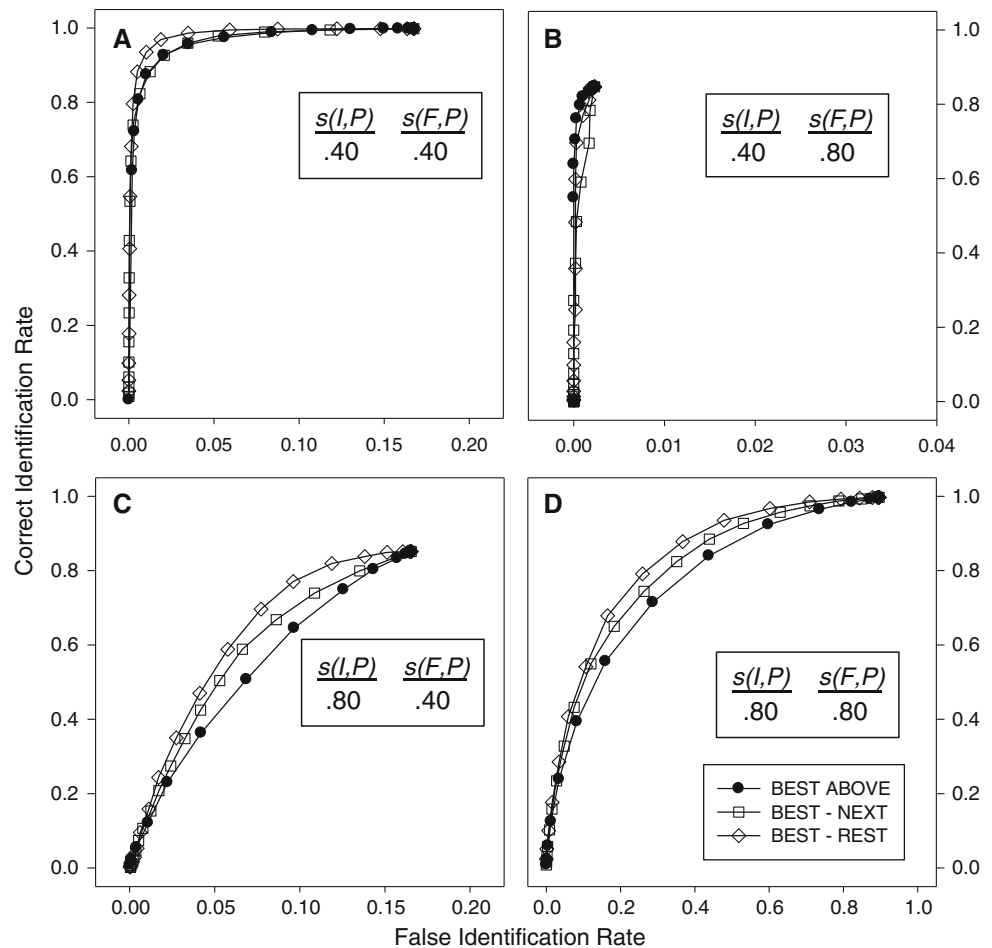
Although the decision rules are based on different kinds of information (absolute versus relative match values), they are in many ways quite similar. For all three models, an identification is made by comparing BEST to some value. For the Best Above Criterion model, the match of the best-matching lineup member is compared to a decision criterion c . For the BEST-NEXT model, the best match is compared to the next-best match, and for the BEST-REST

model, the best match is compared to the average of the match values for the other lineup members. It is important to emphasize that even the absolute decision rule involves a comparison of the best match to some standard. Given this fundamental similarity of the decision rules (and 20/20 hindsight), it is not surprising that performance differences across the models are often quite small.

However, the Best Above Criterion model differs from the two relative judgment models in three ways. First, for the Best Above Criterion model, the value to which BEST is compared, i.e., the decision criterion c , is constant across perpetrator-present and perpetrator-absent lineups. Second, the value of c does not covary with the value of BEST. By contrast, the values of NEXT and REST are lower in perpetrator-absent lineups than they are in perpetrator-present lineups, and those values do covary with the value of BEST. We consider each of these differences in turn.

The Values of NEXT and REST are Lower in Perpetrator-Absent than Perpetrator-Present Lineups. For brevity, the lower value of NEXT and BEST in perpetrator-absent lineups is denoted as the PP-PA difference. The PP-PA difference arises because the

Fig. 3 Suspect ROC curves for same-foils lineups with higher memory accuracy ($a = .6$) and variation in similarity parameters



BEST match tends to be lower for perpetrator-absent lineups than for perpetrator-present lineups. Thus, the foils in the perpetrator-present lineup are competing against a higher best match (often the perpetrator) than are the foils in the perpetrator-absent lineup, which are competing only against other foils. Functionally, this has the effect of lowering the criterion in perpetrator-absent lineups. It is also the case that the PP-PA difference is smaller for REST than for NEXT. The reason for this is that the value of REST is more stable, because it is based on the average of the other five matches, rather than a single next-best match, and because the five foils are the same across perpetrator-present and perpetrator-absent lineups. The decrease in the values of NEXT and REST reduce performance for the BEST-NEXT and BEST-REST models, as the decrease is tantamount to decreasing the decision criterion in perpetrator-absent lineups.

The Values of NEXT and REST Covary with the Value of BEST. The two relative judgment models also differ from the Best Above Criterion model in that the value of NEXT covaries with the value of BEST,

whereas the criterion in the Best Above Criterion model c is constant. This is due to the fact that all lineup members bear some similarity to the perpetrator. Because of the covariance between BEST and NEXT and BEST and REST, the standard deviations of the BEST-NEXT and BEST-REST difference distributions are quite small. As the variances of the BEST-NEXT and BEST-REST distributions decrease, the perpetrator-present and perpetrator-absent lineups overlap less, and performance improves.

These two components affect the behavior of the decision rules in opposite ways: The fact that the values of NEXT and REST are lower in perpetrator-absent lineups than perpetrator-present lineups reduces performance in the BEST-NEXT and BEST-REST models relative to performance in the Best Above Criterion model. By contrast, as the covariation between BEST, NEXT, and REST increase, the performance of the BEST-NEXT and BEST-REST models improves relative to the Best Above Criterion model.

Why does the relative performance of the models depend on the accuracy of the memory trace? Simply, an inaccurate memory is a noisy memory, and this noise reduces the BEST-NEXT and BEST-REST covariances.

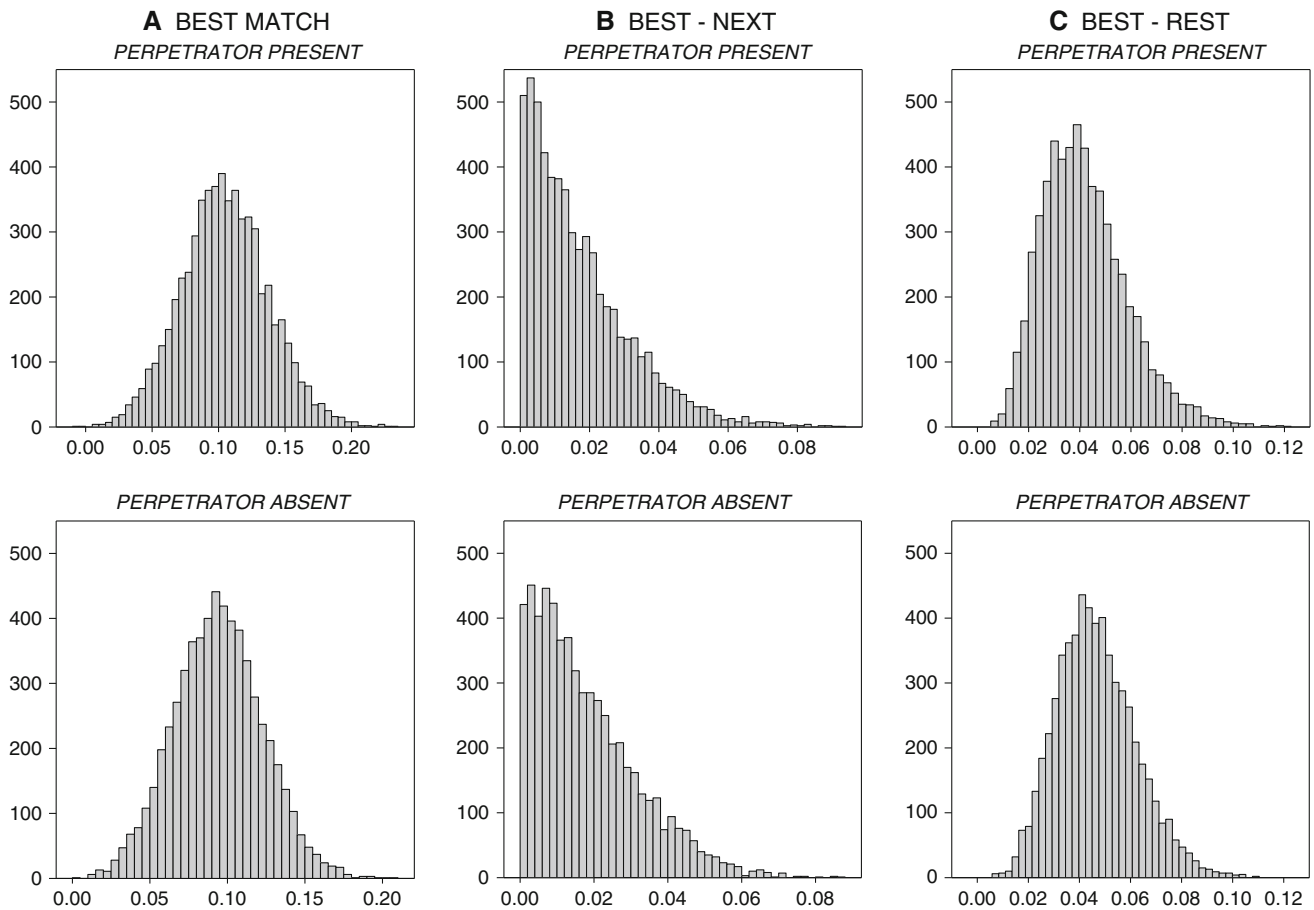


Fig. 4 Distributions for BEST (A), BEST-NEXT (B), and BEST-REST (C) for perpetrator-present and perpetrator-absent, same-foils, lineup

The Underlying Distributions are Very Different. The shapes of the ROC curves depend on the shapes of the underlying distributions of the BEST match, the BEST-NEXT difference, and the BEST-REST difference. An example of these distributions, with the parameters used in the baseline (Fig. 1) simulations, is shown in Fig. 4. It is clear that the distributions for BEST (Panel A) are roughly symmetric and normal in appearance, whereas the distributions for the BEST-REST difference (Panel C) are slightly skewed, and the distributions for the BEST-NEXT difference (Panel B) are very skewed, especially for perpetrator-absent lineups. The shapes of the distributions, and the positive skew of perpetrator-absent lineups in particular, produce a disadvantage for the BEST-NEXT and BEST-REST models with high criterion values as well as the slight advantage for the BEST-REST model at the low criterion values.

Implications of the Simulation Results

Two main points arise from these theoretical analyses:

1. These simulation results challenge the widely held view that eyewitness identification decisions that are the product of an absolute decision strategy are more accurate than identification decisions that are the product of a relative decision strategy. The simulations produced all three patterns of results—an advantage for the absolute decision rule, a disadvantage for the absolute decision rule, and no differences between absolute and relative decision rules. In addition, our analysis of these different patterns of results provides insight regarding the conditions under which these different patterns are likely to arise.
2. Measures of probative value of a suspect identification are not independent of the decision criterion. The specific patterns depend on how probative value is measured. The correct/false ID ratio that is commonly used is particularly sensitive to variation in the decision criterion whereas other measures of probative value were less sensitive to criterion placement. Measured as a conditional probability, i.e., the probability that the suspect is guilty given that the suspect

was identified, probative value estimates increased slightly as criteria were shifted upward. However, small decreases in probative value, measured as a conditional probability of guilt given a suspect identification, are not trivial. As Clark and Godfrey (2009) note, the complementary probability, i.e., the probability that the identified suspect is innocent, can be substantial even when the conditional probability of guilt changes very little. For example, in Table 1, as the conditional probability of guilt increases from .82 to .89 with the increasing criterion, the conditional probability of innocence decreases from .18 to .11, a decrease of nearly 39%. Even d' , which by definition should be invariant across criterion placement, showed some sensitivity to criterion shifts, particularly for relative judgment models. The reason for this is clear from Fig. 4. The underlying distributions are decidedly not normal.

These patterns of results contrast with the intuition that probative value should be invariant of decision criterion. Results showing proportional changes in correct and false identification rates, with no change in the probative value of a suspect identification, are often interpreted as the product of a criterion shift. Conversely, disproportional changes in correct and false identification rates are sometimes interpreted as evidence of a shift in decision strategies. These interpretations both rest on an assumption that is not supported by the simulation results.

Perpetrator- Versus Description-Matched Lineups

As noted earlier, the perpetrator-present and perpetrator-absent lineups will contain the same foils if they are selected based on their match to a photograph of the perpetrator or if they are selected based on their match to a description of the perpetrator. Both methods have been used to create experimental lineups. We have treated both of these cases as examples of the same-foils design, and indeed most eyewitness identification experiments utilize the same-foils design. However, it is reasonable to consider that these two roads to a same-foils lineup design differ—at least in degree. Specifically, because a photograph of the perpetrator has more available information than a description of the perpetrator, the similarity standard for selecting foils may be higher when they are selected based on their resemblance to a photograph rather than their match to a description (Luus & Wells, 1991; Wells, Rydell, & Seelau, 1993). This difference would be captured by the model's foil similarity parameter, $s(F,P)$. To the extent that is the case, then matching to a photo of the perpetrator might produce results akin to those in panels B and D of Figs. 3 and 4 ($s(F,P) = .8$), whereas matching to a verbal

description might produce results more akin to those in panels A and C of Figs. 3 and 4 ($s(F,P) = .2, .4$). Assuming less accurate memory, lineups created with foils selected based on their match to a description of the perpetrator would produce an advantage for the Best Above Criterion decision rule, whereas lineups created with foils selected based on their match to a photograph of the perpetrator would not. There are two caveats regarding this prediction. First, a point that we expand upon later, the correspondence among model parameters, predictions, and experimental tests is not straightforward. Second, differences between matching to a photo versus a description may not simply be a matter of how much information is utilized, but may also reflect different kinds of information. Specifically, descriptions may be limited to information that can be verbalized, a limitation that photographs obviously do not share.

Suspect-Matched Simulations

Simulations of suspect-matched lineups were conducted in similar fashion to those conducted for same-foils lineups. However, the results were much simpler, and thus are summarized simply in Figs. 5 and 6. Figure 5 used the same parameters used in the baseline for same-foils lineups, the results of which were shown in Fig. 1. Figure 6 shows the variation in the similarity parameters with

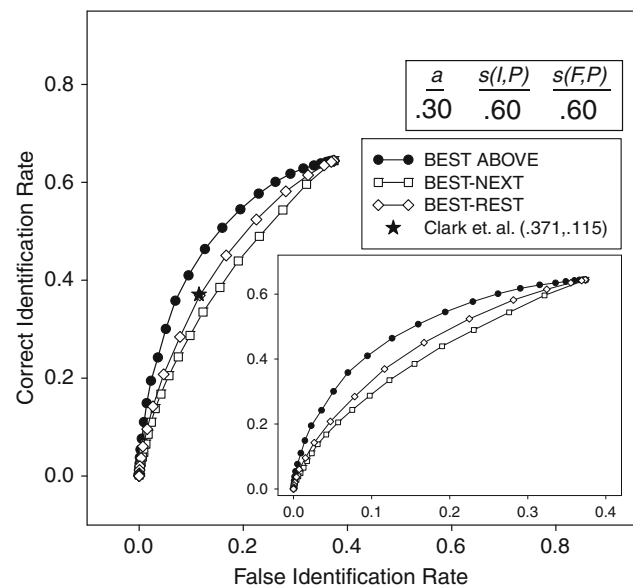
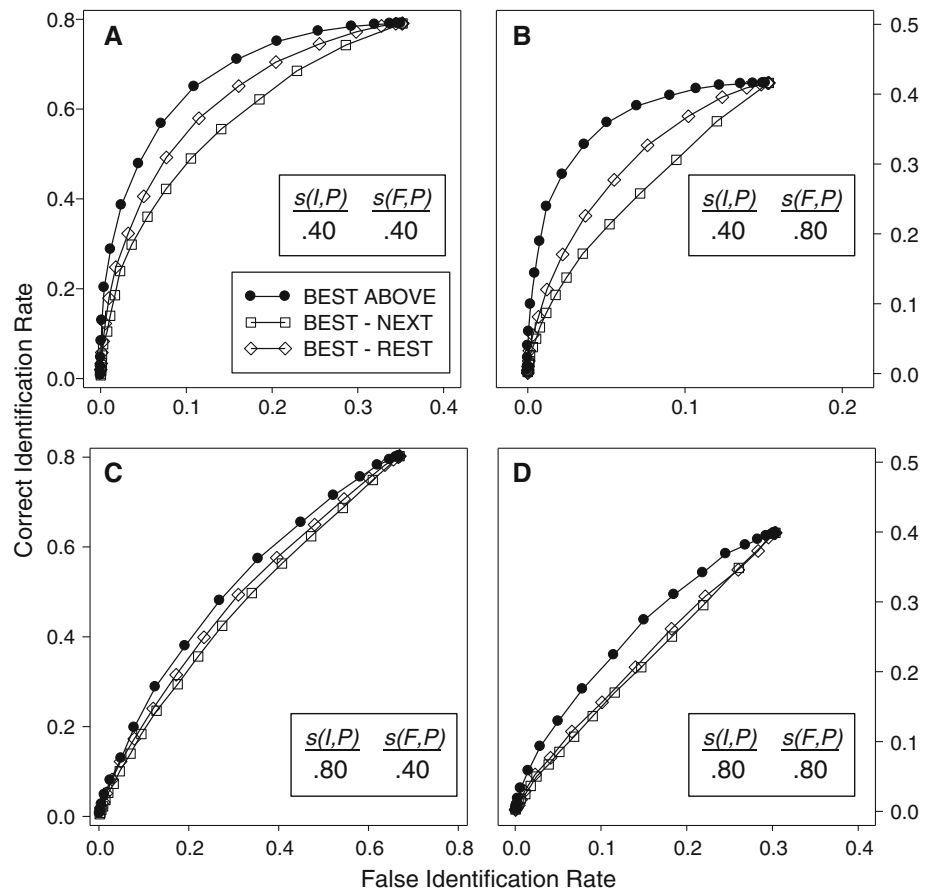


Fig. 5 Suspect ROC curves for suspect-matched lineups with baseline parameters, $a = .3$, $s(I,P) = .6$, and $s(F,P) = .6$, identical to those in Fig. 1. The star symbol represents typical results as the average correct (.371) and false identification (.115) rates from Clark et al. (2008)

Fig. 6 Suspect ROC curves for suspect-matched lineups, with variation in similarity parameters with memory accuracy, $a = .3$



$a = .3$. (The pattern of results with other values of a did not change.)

Baseline

The same parameter values used for the same-foils baseline were also used for the suspect-matched lineup baseline, and yielded results very close to typical performance for suspect-matched lineups, based on the average correct and false identification rates from the meta-analysis by Clark et al. (2008). Although it is tangential to the goals of this article, it is worth noting that the model, with the same parameters, produced ROC curves that included the average correct and false identification rates from the Clark et al. (2008) meta-analysis for both same-foils (Fig. 1) and suspect-matched foils (Fig. 5) designs.

The results of the baseline simulation show the highest level of performance for the Best Above Criterion Model, followed by the BEST-REST model, with the lowest level of performance shown for the BEST-NEXT model. In contrast to the same-foils baseline simulation in which the Best Above Criterion model was only slightly better than the relative judgment models, Fig. 5 shows that the advantage of the Best Above Criterion model over the two

relative judgment models is substantial. Before discussing the basis or implications of these results, we first consider their generality.

The three main parameters of the model, a , $s(I,P)$, and $s(F,S)$, were varied as they were for simulations of same-foils lineups. However, because the results were very consistent, only the results for $a = .3$ are shown in Fig. 6. The ordering of the ROC curves was invariant: Best Above Criterion > BEST-REST > BEST-NEXT. When the similarity parameters were both set to .8, the BEST-REST and BEST-NEXT models converge (Panel D); however, this is simply a floor effect as performance approached chance-level discrimination between the guilty and the innocent.

What produces this consistent pattern of results? Note that in a suspect-matched lineup, the foils in the perpetrator-absent lineup are selected in the same way they are selected in the perpetrator-present lineup. More to the point, if the parameter $s(F,S)$ is .40 in the perpetrator-present lineup then it is .40 in the perpetrator-absent lineup. Consequently, the difference between the best match and the other lineup members is almost as large for perpetrator-absent lineups as it is for perpetrator-present lineups. This reduces the utility of the BEST-NEXT and BEST-REST differences for distinguishing between guilt and innocence.

Also, the values of NEXT and REST drop much more for suspect-matched lineups than they do for same-foils lineups. The performance of the BEST-REST model is better than that of the BEST-NEXT model because the value of NEXT decreases more for perpetrator-absent lineups than does the value of REST.

Implications

According to a survey of police officers, conducted by Wogalter et al. (2004), police officers typically create lineups by selecting foils that look similar to the suspect. For these suspect-matched lineups, the suspect ROC curves showed a clear and consistent superiority for the Best Above Criterion Model over the two Relative Difference Models. Put another way, a suspect identification generated by a Best Above Criterion decision strategy should be more diagnostic of the suspect's guilt than a suspect identification generated by a Relative Difference decision strategy, and this advantage for the Best Above Criterion decision rule extends beyond what would be expected by a conservative shift in a decision criterion. Conversely, a suspect identification based on a relative judgment from a suspect-matched lineup has a greater risk of being a false identification than a suspect identification based on an absolute judgment.

These results are entirely consistent with the view that absolute judgment strategies lead to more accurate eyewitness identification decisions than relative judgment strategies. Moreover, the superiority of absolute judgments was observed with lineups that are compositionally similar to those that are typically used in actual criminal investigations.

General Discussion

It is a widely held view that eyewitnesses' reliance on relative judgments in the course of making identification decisions reduces the accuracy of those decisions and increases a source of error in the criminal justice system. This view is widely held not only within the research community, but also by policy-making groups (Commonwealth of Virginia, 2005; State of Wisconsin, 2005) and advisors (Schuster, 2007).

The simulation results reported here were consistent with that widely held view for lineups in which foil selection is based on similarity to the suspect. However, for same-foils lineups—those that are typically used in eyewitness identification experiments—the simulations did not show a consistent advantage for absolute over relative decision rules.

In the remainder of the article we discuss (1) the basis of the model's behavior, (2) the implications of the simulation results for the criminal justice system and experimental research on eyewitness identification, and (3) the potential limitations of this study, with a look forward to future research.

Basis of the Model's Behavior

The behavior of the three (excluding the One Above Criterion Model) decision rules depends on the shapes of the underlying distributions, which can be decomposed into three components: (1) the decrease in the value of the comparison indices in perpetrator-absent lineups relative to perpetrator-present lineups, (2) the covariance among BEST, NEXT, and REST, and (3) the skew of the BEST-NEXT distribution. Each of these components is discussed briefly below.

The value of the best match is lower in a perpetrator-absent lineup than in a perpetrator-present lineup. The reason, of course, is that the expected match of the perpetrator in the lineup to the perpetrator in memory is higher than the expected match of any other lineup member. Consequently, the values of NEXT and REST are also lower in perpetrator-absent lineups than in perpetrator-present lineups. These values are critical in the decision-making in the BEST-NEXT and BEST-REST models, and a decrease in their values is tantamount to a decrease in the decision criterion. For the Best Above Criterion model, however, the criterion is constant across perpetrator-present and perpetrator-absent lineups. This gives an advantage to the Best Above Criterion Model over the other two models.

The second factor is that the values of BEST, NEXT, and REST are correlated. This correlation produces a decrease in the variances of the BEST-NEXT and BEST-REST distributions. This factor facilitates performance for the BEST-NEXT and BEST-REST models relative to the Best Above Criterion model.

The third factor is peculiar to the distribution of BEST-NEXT differences. This distribution is somewhat skewed for perpetrator-present lineups and extremely skewed for perpetrator-absent lineups. The consequence of this skew is that false identifications are higher for high criterion values than they would be without the skew, and thus the slope of the ROC curve is flattened somewhat as it rises from the (0,0) point.

For same-foils lineups these three factors combine to produce a wide range of outcomes: an advantage for the Best Above Criterion model when the memory trace was less accurate and foil similarity high, an advantage for the BEST-REST and BEST-NEXT models when the memory trace was more accurate, particularly when the innocent

suspect was more similar to the perpetrator; and many cases in which the differences between models were so small as to be nearly indistinguishable. For suspect-matched lineups, however, the first factor—the drop in the values of NEXT and REST—is much larger (because the foils in perpetrator-absent lineups are less similar to the perpetrator than the foils in perpetrator-present lineups). As a consequence, this factor alone determines the relative performance of the models and produces a consistent advantage for the Best Above Criterion model.

Implications for the Criminal Justice System

The most important implication of the absolute-relative distinction, as it was initially framed by Wells (1984), is that witnesses' reliance on relative judgments undermines the reliability of the identification evidence, and increases the relative risk of a false identification that can ultimately lead to a wrongful conviction.

This concern still stands. Although our analyses showed a pattern of small and inconsistent results for same-foils lineups, the analyses could not have provided a stronger confirmation regarding the superiority of absolute judgments for suspect-matched lineups. It cannot be overemphasized that suspect-matched lineups are the lineups typically used in real criminal investigations. Thus, for the real-world circumstances to which eyewitness identification research is applied, the present results add to the concerns regarding witnesses' use of relative judgments.

Implications for Eyewitness Identification Research

Although the concern about relative judgments still stands, there is much empirical work to be done, as the relevant experiments have all been conducted with same-foils lineups, for which the predictions are the least straightforward. None of the relevant experiments have been conducted with suspect-matched lineups, the condition for which the superiority of the Best Above Criterion model is predicted most clearly.

Clark et al. (2008) and Clark and Tunnicliff (2001) have argued that the same-foils design is an artifact specific to lineups as they are constructed for experiments, and that suspect-matched and same-foils designs produce different patterns of results. The simulations reported here add to those concerns regarding the use of same-foils lineups in eyewitness identification experiments. The conclusions one draws may be quite different depending on whether the experiment uses suspect-matched or same-foils lineups.

Experimental results showing changes in the probative value of a suspect identification have been taken as evidence of a shift between absolute and relative decision strategies. Conversely, experimental results showing

changes in both correct and false identification rates, with no change in the probative value of a suspect identification, have been taken as evidence against such a shift. The present analyses suggest that this reasoning is limited at best. Shifts from relative to absolute decision rules may under some conditions produce no change in the probative value of a suspect identification, whereas criterion shifts within a single model will. Thus, research into the decision processes underlying eyewitness identification decisions must be re-evaluated.

Noting again that all of the relevant experiments have been conducted with same-foils lineups, we must reconsider the implications of results that show no change in the probative value of a suspect identification across conditions. These results are consistent with the view that witnesses shift between absolute and relative decision strategies. Consider next those cases in which the probative value of a suspect identification does vary across conditions. It has generally been assumed that such a result cannot be produced by a criterion shift. Our results show that the probative value of a suspect identification increases with the decision criterion. There are limitations to this, of course. For example, Lindsay and Wells (1985) showed that the ratio of correct to false identifications increased from 1.35 for simultaneous lineups to 3.06 for sequential lineups. Such a sizeable increase in probative value may not be within the range of a criterion shift, and some other mechanism may be necessary to produce such a large difference (see Goodsell et al., 2010 regarding possible other mechanisms).

There are testable predictions that arise from these simulation results. We attach a word of caution, however, as the link between simulation and experiment is likely to be complicated by a host of factors. We discuss the predictions first, followed by the caveats.

The Predictions

The identification of a suspect should have higher probative value for witnesses utilizing a Best Above Criterion decision rule than for witnesses using a BEST-NEXT or BEST-REST decision rule, provided that the following conditions hold.

The lineups are created with foils selected based on their match to the suspect.

The lineups are created with foils that are the same for perpetrator-present and perpetrator-absent lineups, provided that the following additional conditions apply: (a) memory is relatively inaccurate, and (b) the foils have high similarity to the perpetrator.

The identification of a suspect should have higher probative value if it arises from a relative judgment strategy (with the clearest predictions for the BEST-REST model,

in particular), compared to a Best Above Criterion rule, if the following condition holds:

Perpetrator-present and perpetrator-absent lineups contain the same foils and (a) the witness's memory is relatively accurate, and (b) the innocent suspect is highly similar to the perpetrator.

The Caveats to the Predictions

These “straightforward” predictions come with a number of caveats, however.

1. The probative value of a suspect identification will also increase if the witness simply adopts a more conservative decision criterion. Thus, the difference in probative value due to a shift in strategy versus a shift in criterion (with no change of strategy), according to the present analyses, is a matter of degree. Moreover, the advantage of an absolute judgment strategy over a relative judgment strategy depends on the specific relative judgment strategy that is being considered. The advantage of the Best Above Criterion model over the BEST-REST model, while consistent, is not as large as the advantage of the Best Above Criterion Model over the BEST-NEXT model. Consequently, observing a change in the probative value of a suspect identification across conditions of an experiment does not, by itself, provide a clear test between the various models.
2. Any empirical test is further complicated by the fact that the experimental manipulations are unlikely to produce pure measures of the underlying decision processes. For example, the claim has been made that biased instructions induce witnesses to shift from absolute to relative judgments. However, it is unlikely that *all* of the witnesses in the unbiased instruction condition would utilize absolute judgments and that *all* of the witnesses in the biased instruction condition would utilize relative judgments. Thus, a small observed change in probative value across conditions is consistent with a simple criterion shift model, but is also consistent with the highly likely possibility that the experimental manipulations are less than 100% effective (i.e., all witnesses in the unbiased instructions condition using absolute judgments and all witnesses in the biased instructions condition using relative judgments).
3. The linkage between model parameters and experimental conditions is not straightforward. Some of the predictions outlined above depend on special conditions, i.e., the accuracy of memory and the similarity relationships among lineup members. If a prediction is not confirmed in an experimental test, the question will

arise as to whether the necessary condition was properly instantiated. For example, if the Best Above Criterion Model does not show an advantage in an experimental test with same-foils lineups, one may question whether the accuracy of memory, which needs to be low in order to produce the advantage, was low enough. The same issue arises regarding the similarity of the foils and of the innocent suspect.

This may seem a rather pessimistic appraisal of the ability to test between models with experimental data. Quite to the contrary, the present theoretical analyses should play a very important role in future empirical research by clarifying the predictions that models make, as well as the predictions they do not make. The point in listing the above complications and caveats is that the experimental tests will probably not be simple or straightforward (even if the predictions seem straightforward), and it is unlikely that much clarity will emerge from only one or two experiments.

With all of the caveats duly noted, one prediction is particularly striking. Experimental manipulations that induce witnesses to switch between absolute and relative decision rules should produce larger and more consistent results in suspect-matched lineups than in same-foils lineups. Thus, for example, to the extent that simultaneous lineups bring out witnesses' tendencies to rely on relative judgments, and sequential lineups reduce that reliance on relative judgments, the sequential lineup advantage should be larger in suspect-matched lineups than in same-foils lineups. The same logic holds regarding biased and unbiased lineup instructions. Any probative-value advantage for unbiased instructions should be larger in suspect-matched lineups than in same-foils lineups.

Limitations, Other Possible Models, and Future Theoretical Development

One potential limitation of the present work is that it focuses on a particular implementation of absolute and relative judgments and is tied to a particular model, i.e., the WITNESS model. However, this criticism is weakened by the fact that two different versions of each model were considered. One of the absolute judgment models was easily rejected, and the differences between the two relative judgment models were a matter of degree, and a very small degree indeed for same-foils lineups.

Another potential limitation in this study concerns the extent to which the conclusions are specific to the WITNESS model. The WITNESS model is a very generic version of a class of models for recognition memory called global matching models (see Clark & Gronlund, 1996, for a review). Like other models (Hintzman, 1988; Murdock,

1982; Pike, 1984), it represents information as vectors of feature-elements, it represents similarity as feature overlap, and it computes similarity by a dot product comparing a test item vector (lineup member) to a memory vector. These fundamental similarities across models suggest that the results obtained here should generalize across other models of this ilk. However, intuition can be an unreliable tool for generating predictions of other models. It would be useful to see how absolute and relative decision rules behave in other models.

Finally, the simulations reported here compared decision rules as they operate in pure form in simultaneous lineups. These simulations are relevant in the comparison of simultaneous and sequential lineups to the extent that simultaneous-sequential differences arise from witnesses employing different decision rules for simultaneous versus sequential lineups. Although the dominant explanation regarding simultaneous-sequential differences is based on decision processes (Lindsay & Wells, 1985), other factors may also play a role. For example, simultaneous-sequential differences may depend in part on the position of the suspect in the sequential lineup (Carlson, Gronlund, & Clark, 2008; Clark & Davey, 2005; Goodsell et al., 2010; Gronlund, Carlson, Dailey, & Goodsell, 2009; McQuiston-Surrett, Malpass, & Tredoux, 2006). The present simulations provide a baseline for considering the operation of relative and absolute judgments in simultaneous and sequential lineups; however, a wide range of other memory and decision processes may be relevant as well.

Our implementation of absolute and relative judgment strategies is consistent with the language Wells used to describe them in his 1984 paper, as well as the usage of the terms since the 1984 paper. However, the connection between concept and implementation is never 1-to-1. There is more than one way to implement the same idea. There are other ways of instantiating relative judgments that should also be explored. For example, Dunning and Stern (1994) suggested that witnesses may arrive at identification decisions as a process of elimination. One possibility, suggested by Clark, Marshall, and Rosenthal (2009) and inspired by the Elimination by Aspects model of Tversky (1977), is that witnesses eliminate lineup members based on different feature mismatches. For example, lineup member A might be eliminated because of a mismatch on feature *x*, whereas lineup member B is eliminated based on a mismatch on feature *y*. The performance of this version of a relative judgment model would depend in very large part on how attention shifted across dimensions and across elimination decisions. Alternatively, absolute and relative decision rules may combine in a variety of mixed models that vary the decision processes both within and across witnesses. The exploration and development of other models, and the possibilities for new empirical research

described earlier, are promising avenues for future research.

Acknowledgments The authors are indebted to Scott Gronlund, Brian Cutler, Margaret Kovera, Gary Wells, and the many anonymous reviewers who slogged through various versions of this manuscript to make this a much better paper. The research was supported by the National Science Foundation, Law and Social Sciences Grant SES 0647947.

References

- Brandon, R., & Davies, C. (1973). *Wrongful imprisonment: Mistaken convictions and their consequences*. Hamden, CT: Archon Books.
- Breneman, J. S., & Clark, S. E. (2008). *Probative value of absolute and relative decision rules*. Paper presented at the 19th Annual Meeting of the American Psychology—Law Society, Jacksonville, FL.
- Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *14*, 118–128. doi:10.1037/1076-898X.14.2.118.
- Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 739–755. doi:10.1037/0278-7393.30.4.739.
- Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629–654. doi:10.1002/acp.891.
- Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior*, *29*, 151–172. doi:10.1007/s10979-005-2418-7.
- Clark, S. E., & Godfrey, R. (2009). Innocence risk and the probative value of eyewitness identification evidence. *Psychonomic Bulletin & Review*, *16*, 22–42. doi:10.3758/PBR.16.1.22.
- 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. doi:10.1007/s10979-006-9082-4.
- Clark, S. E., Marshall, T. E., & Rosenthal, R. (2009). Lineup administrator influences on eyewitness identification decisions. *Journal of Experimental Psychology: Applied*, *15*, 63–75. doi:10.1037/a0015185.
- 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. doi:10.1023/A:1005463020252.
- Commonwealth of Virginia. (2005). *Mistaken eyewitness identification: Report of the Virginia State Crime Commission*.
- 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. doi:10.1037/0022-3514.67.5.818.
- Dysart, J. E., & Lindsay, R. C. L. (2007). Show-up identifications: Suggestive technique or reliable method? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, vol. II: Memory for people* (pp. 137–153). Mahwah, NJ: Erlbaum.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669. doi:10.1037/0033-295X.103.4.650.

- Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior*. doi:10.1007/s10979-009-9215-7.
- Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology*, 19, 23–37. doi:10.1002/acp.1047.
- Gronlund, S. D., Carlson, C. A., Dailey, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, 15, 140–152. doi:10.1037/a0015082.
- Gross, S. R., Jacoby, K., Matheson, D. J., Montgomery, N., & Patil, S. (2005). Exonerations in the United States 1989 through 2003. *Journal of Criminal Law and Criminology*, 95, 523–560.
- Hastie, R., & Stasser, G. (2000). Computer simulation methods for social psychology. In: H. T. Reis & C. M. Judd (Eds.). *Handbook of research in social and personality psychology* (pp. 85–114). New York: Cambridge University Press.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551. doi:10.1037/0033-295X.95.4.528.
- Huff, C. R., Rattner, A., & Sagarin, E. (1996). *Convicted but innocent*. Thousand Oaks, CA: Sage.
- Kaye, D. H. (1986). Quantifying probative value. *Boston University Law Review*, 66, 761–766.
- Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology*, 15, 659–671. doi:10.1002/acp.739.
- Lindsay, R. C. L. (1999). Applying applied research: Selling the sequential lineup. *Applied Cognitive Psychology*, 13, 219–225. doi:10.1002/(SICI)1099-0720(199906)13:3<219:AID-ACP562>3.0.CO;2-H.
- Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology*, 84, 315–321. doi:10.1037/0021-9010.84.3.315.
- Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology*, 76, 741–745. doi:10.1037/0021-9010.76.5.741.
- Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship between lineup fairness and identification accuracy. *Law and Human Behavior*, 4, 303–314. doi:10.1007/BF01040622.
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presentations. *Journal of Applied Psychology*, 70, 556–564. doi:10.1037/0021-9010.70.3.556.
- Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distractors for lineups. *Law and Human Behavior*, 15, 43–57. doi:10.1007/BF01044829.
- MacLin, O. H., Meissner, C. A., & Zimmerman, L. A. (2005). PC eyewitness: A computerized framework for the administration and practical application of research in eyewitness psychology. *Behavior Research Methods*, 37, 324–334.
- Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, 66, 482–489. doi:10.1037/0021-9010.66.4.482.
- McQuiston-Surrett, D., Malpass, R. S., & Tredoux, C. G. (2006). Sequential vs. simultaneous lineups: A review of methods, data, and theory. *Psychology, Public Policy, and Law*, 12, 137–169. doi:10.1037/1076-8971.12.2.137.
- Meissner, C. A., Tredoux, C. G., Parker, J. F., & MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory and Cognition*, 33, 783–792.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609–626. doi:10.1037/0033-295X.89.6.609.
- Ogloff, J. R. P. (2000). Two steps forward and one step backward: The law and psychology movement(s) in the 20th century. *Law and Human Behavior*, 24, 457–483.
- Penrod, S., & Hastie, R. (1980). A computer simulation of jury decision making. *Psychological Review*, 87, 133–159. doi:10.1037/0033-295X.87.2.133.
- Pike, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review*, 91, 281–294. doi:10.1037/0033-295X.91.3.281.
- Pozzulo, J. D., Crescini, C., & Lemieux, J. M. T. (2008). Are accurate witnesses more likely to make absolute judgments? *International Journal of Law and Psychiatry*, 31, 495–501. doi:10.1016/j.ijlp.2008.09.006.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134. doi:10.1037/0033-295X.88.2.93.
- Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, 137, 528–547. doi:10.1037/a0012712.
- Scheck, B., Neufeld, P., & Dwyer, J. (2000). *Actual innocence: Five days to execution and other dispatches from the wrongly convicted*. New York: Doubleday.
- Schuster, B. (2007). Police lineups: Making eyewitness identification more reliable. *National Institute of Justice Journal*, 258, 2–9.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4, 145–166.
- Small, M. A. (1993). Legal psychology and therapeutic jurisprudence. *St. Louis University Law Journal*, 37, 675–713.
- State of Wisconsin. (2005). *Model policy and procedure for eyewitness identification*.
- Stebly, 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. doi:10.1023/A:1012888715007.
- Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1, 1–26. doi:10.1111/1529-1006.001.
- Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401–409. doi:10.1037/h0058700.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352. doi:10.1037/0033-295X.84.4.327.
- Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, 14, 89–103. doi:10.1111/j.1559-1816.1984.tb02223.x.
- Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, 48, 553–571. doi:10.1037/0003-066X.48.5.553.
- Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, 78, 835–844. doi:10.1007/BF01044829.
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, 23, 603–647. doi:10.1023/A:1025750605807.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114, 152–176. doi:10.1037/0033-295X.114.1.152.
- Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of U.S. police on preparation and conduct of lineups. *Psychology, Crime & Law*, 10, 69–82. doi:10.1080/10683160410001641873.