Law and Human Behavior

New Signal Detection Theory-Based Framework for Eyewitness Performance in Lineups

Jungwon Lee and Steven D. Penrod

Online First Publication, August 1, 2019. http://dx.doi.org/10.1037/lhb0000343

American Psychological Association
American Psychology-Law Society

Law and Human Behavior

0147-7307/19/$12.00

http://dx.doi.org/10.1037/lhb0000343

New Signal Detection Theory-Based Framework for Eyewitness Performance in Lineups

Jungwon Lee
John Jay College of Criminal Justice and The Graduate Center, City University of New York

Steven D. Penrod
John Jay College of Criminal Justice, City University of New York

Objectives: Eyewitness research has adapted signal detection theory (SDT) to investigate eyewitness performance. SDT-based measures in yes/no tasks fit well for the measurement of eyewitness performance in show-ups, but not in lineups, because the application of the measures to eyewitness identifications neglects the role of fillers. In the present study, we introduce a SDT-based framework for eyewitness performance in lineups—Multi-d' Model. ***Method:*** The Multi-d' model provides multiple discriminability measures which can be used as parameters to investigate eyewitness performance. We apply the Multi-d' model to issues in eyewitness research, such as the comparison of eyewitness discriminability between show-ups and lineups; the influence of lineup bias on eyewitness performance; filler selection methods (match-to-description vs. match-to-suspect); eyewitness confidence; and lineup presentation modes (simultaneous vs. sequential lineups). ***Results:*** The Multi-d' model demonstrates that the discriminability of a guilty suspect from an innocent suspect is a function of discriminability involving fillers; and underscores that the decisions that eyewitnesses make in lineups can be regarded from two perspective—detection and identification. ***Conclusions:*** We propose that the Multi-d' model is a useful tool to understand decisionmakers' performance in a variety of compound decision tasks, as well as eyewitness identifications in lineups.

Public Significance Statement
This study introduces multi-d' model, which is a framework for explaining eyewitness performance in lineups based on signal detection theory. The multi-d' model calls attention to the role of fillers for constructing eyewitnesses' ability to correctly identify a guilty suspect while avoiding misidentifications of an innocent suspect. The multi-d' model also provides multiple measures which can be used as parameters to investigate eyewitness performance.

Keywords: signal detection theory, eyewitness identification, lineup

Supplemental materials: http://dx.doi.org/10.1037/lhb0000343.supp

Eyewitness research has adapted signal detection theory (SDT) to measure eyewitness performance (Meissner, Tredoux, Parker, & MacLin, 2005; Palmer & Brewer, 2012; Wixted & Mickes, 2014). SDT explains human performance in tasks which involve the discrimination of stimulus-presence versus stimulus-absence. For example, a participant participates in a series of yes/no trials in which the participant listens for a signal or a noise and must respond "Yes, it's a signal" when he or she hears a signal or "No, it's not a signal" when he or she hears only noise. That is, the participant must discriminate signal-presence from signal-absence. The participant's response can be classified into a 2×2 matrix (see Table 1). The participant decides whether to respond affirmatively or negatively based on stimulus strength. If the strength of a stimulus is greater than a decision criterion set by the participant, then he or she would respond affirmatively. Otherwise, he or she would respond negatively. When the participant sets a stringent criterion, the degree of stimulus strength needed for an affirmative response is greater than the requirements within a loose criterion. Figure 1 shows distributions of signal strength and noise strength with the placement of a decision criterion which must be exceeded in order for the participant to indicate he or she has detected a signal. As shown in Figure 1, when the criterion is stringent, the participant is less likely to respond affirmatively, which

Jungwon Lee, Department of Psychology, John Jay College of Criminal Justice, and The Graduate Center, City University of New York; Steven D. Penrod, Department of Psychology, John Jay College of Criminal Justice, City University of New York.

Portions of the present study were presented at the 2019 meeting of the Society for Applied Research in Memory and Cognition, Cape Cod, Massachusetts. The present study is based upon work supported by the National Science Foundation under Grant SES-1754079. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this article should be addressed to Jungwon Lee, Department of Psychology, John Jay College of Criminal Justice, City University of New York, 524 West 59th Street, New York, NY 10019. E-mail: jlee6@gradcenter.cuny.edu

reduces both mistaken identifications of a noise as the signal (i.e., false alarm) and correct identifications of the signal (i.e., hit). On the other hand, when the participant uses a loose criterion, he or she is more likely to respond affirmatively, which increases both false alarms and hits.

Parameters have been introduced to measure the ability to discriminate between signal-presence and signal-absence (e.g., A', Pollack & Norman, 1964; A_z, Swets & Pickett, 1982; d' and area under the curve, Green & Swets, 1966). To measure discriminability of a signal from a noise in yes/no tasks, memory studies have frequently used indices that estimate the distance between the mean of the signal strength distribution and the mean of the noise strength distribution. Equations for computing the distance vary with assumptions regarding variances of the signal and noise distributions. When different variances are assumed between signal and noise distributions ($\sigma^2_{signal} \neq \sigma^2_{noise}$), d_a and d'_e (Green & Swets, 1966; Irwin & McCarthy, 2013) can be used to estimate the distance between the signal and noise distributions.

$$d_a = \frac{\mu_{signal} - \mu_{noise}}{\sqrt{\frac{(\sigma^2_{signal} + \sigma^2_{noise})}{2}}} \qquad d'_e = \frac{\mu_{signal} - \mu_{noise}}{\frac{(\sigma_{signal} + \sigma_{noise})}{2}}$$

When equal variance is assumed between signal and noise distributions ($\sigma^2_{signal} = \sigma^2_{noise}$), the equations for d_a and d'_e reduce to the equation for d' below; and the distance between the two distributions is estimated in standard deviation units of the noise distribution. d' can be also calculated with the inverse cumulative distribution function of hit and false alarm rates, as in the following equation; d' = z(hit rate) − z(false alarm rate).[1]

$$d' = \frac{\mu_{signal} - \mu_{noise}}{\sigma_{noise}}$$

Because basic memory recognition tasks generally demonstrate unequal variance between signal and noise distributions (Mickes, Wixted, & Wais, 2007; Starns & Ratcliff, 2014), d' computed from the z-transformed hit and false alarm rates may not precisely estimate eyewitness discriminability under the unequal variance assumption. Nevertheless, in addition to ease of computation, we use the d' as a primary index for the multi-d' model, for the following three reasons.

First, unlike basic memory recognition tasks which demonstrate $\sigma^2_{signal} \neq \sigma^2_{noise}$, eyewitness identification data from lineup tasks often supports equal variance models in model-recovery simulations. For example, when performing model-recovery simulations with empirical eyewitness data, unequal variance models (vs. equal variance models) did not produce better model fits or even produced worse model fits (Colloff, Wade, & Strange, 2016; Wixted, Vul, Mickes, & Wilson, 2018).

Table 1
A 2 × 2 Matrix of Participant Performance in Yes/No Signal Detection Tasks

	Actual status	
Participant response	Presence of signal	Absence of signal
"Yes, it's a signal"	True positive	False positive
"No, it's not a signal"	False negative	True negative

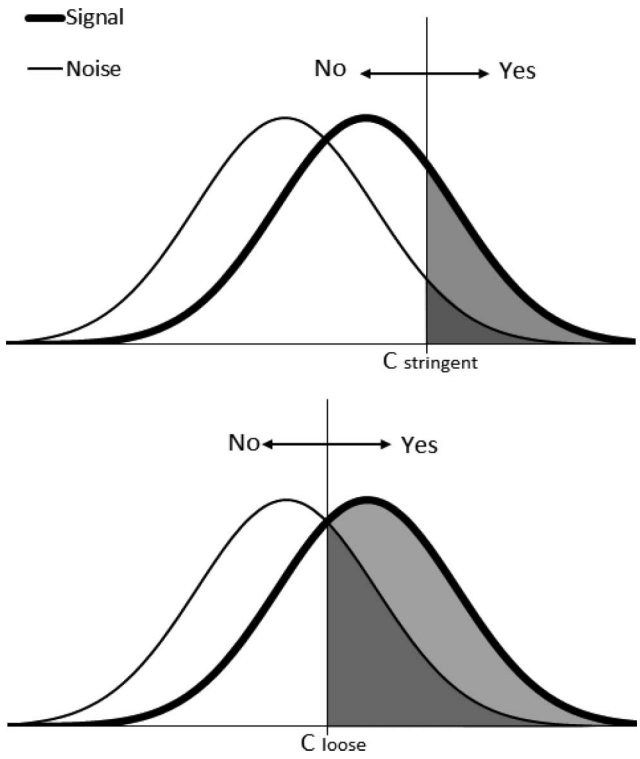


Figure 1. Criterion shift and response accuracy in SDT.

Second, the difference in the value of estimated discriminability between equal variance and unequal variance assumptions could be minor. To the best our knowledge, Wixted and Mickes' (2018) study is currently the only published eyewitness study which provided values of theoretical discriminability under an unequal variance assumption. In their study, using confidence-based ROC data from Mickes et al. (2017), ensemble model estimated the discriminability of a guilty suspect from an innocent suspect as 2.24 under an unequal variance assumption, whereas the z-transformed hit rate minus the z-transformed false alarm rate estimated the discriminability as 2.20. In instruction-based ROC data from Mickes et al. (2017), their ensemble model yielded the discriminability values of 2.06 for the two extreme conditions and 2.22 for the two more neutral conditions, while the z transformation equation produced the discriminability of 1.92 and 2.23, respectively.

Third, d' computed from z-transformed hit and false alarm rates is regarded as an index of *empirical* discriminability of eyewitnesses in lineups, which is a major interest of policymakers (Wixted & Mickes, 2015b, 2018). The implications of the distinction between empirical discriminability and theoretical/underlying discriminability in eyewitness identifications have been discussed in prior eyewitness studies (Wells, Smalarz, & Smith, 2015; Wixted & Mickes, 2015b, 2018). Wixted and Mickes (2018) noted that theoretical discriminability is a major interest of theoreticians who investigate "unobservable memory or perceptual signals from two classes of repeatedly presented stim-

[1] z(p) is the inverse cumulative distribution function of a normal distribution (MacMillan & Creelman, 2005). The formula in Excel program is NORM.S.INV(hit rate) − NORM.S.INV(false alarm rate).

uli" (p. 2). Theoretical discriminability is generally estimated either by simulations or by fitting models to data (Palmer & Brewer, 2012; Smith, Wells, Smalarz, & Lampinen, 2018; Wixted et al., 2018). However, empirical discriminability (e.g., partial area under the ROC curve) is the measure of applied interest, and it can be influenced by additional sources of variance over and above the memory signals that lineup members generate (e.g., variability in criterion placement and other latent variables). Following Mickes, Moreland, Clark, and Wixted's (2014) recommendation, Wixted and Mickes (2018) argued that d' computed from a pair of z-transformed hit and false alarm rates provides a useful proxy for empirical discriminability. Used this way, it is an important index for policymakers who mainly care about "the degree to which participants correctly sort target and foil stimuli into their true categories" (Wixted & Mickes, 2018, p. 2). We similarly argue that z-transformed measures of other aspects of witness performance can provide us with important insights into the manner in which a forensically relevant factors influence multiple aspects of witness performance.

Given that the purpose of the multi-d' model is not to precisely estimate the underlying discriminability of a guilty suspect from an innocent suspect, but to consider issues such as how fillers influence the empirical discriminability of a guilty suspect from an innocent suspect, we use d' computed from z-transformed hit and false alarm rates as the distance measure of memory-strength distributions of lineup members in multi-d' model.

Multiple Discriminability Measures of Eyewitness Identification in Lineups

The traditional framework of SDT in yes/no tasks fits well for the measurement of eyewitness performance in *show-ups*. In parallel with the task of discriminating between signal-presence and signal absence, eyewitnesses viewing show-ups must discriminate whether a perpetrator is present or absent in show-ups (i.e., perpetrator = signal; innocent suspect = noise). However, when applying SDT-based measures in yes/no tasks to eyewitness performance in *lineups*, the application encounters a problem, which is caused by the presence of fillers in lineups. Because lineups include fillers as well as a guilty/innocent suspect, memory-strength distributions of fillers arguably should be also drawn for eyewitnesses viewing lineups. Figure 2 illustrates that four different discriminability measures emerge by considering the memory-strength distribution of fillers—the discriminability of a guilty suspect from an innocent suspect, d'(GI); the discriminability of a guilty suspect from fillers in target-present (TP) lineups, $d'(GF_p)$; the discriminability of an innocent suspect from fillers in target-absent (TA) lineups, $d'(IF_a)$; and the discriminability of fillers in TP lineups from fillers in TA lineups, $d'(F_aF_p)$. The latter three discriminability measures arise in lineups but not in show-ups.

When an innocent suspect does not stand out from fillers in TA lineups (i.e., a perfectly fair TA lineup), the innocent suspect and TA filler distributions will overlap with each other. When the same fillers are used in TP and TA lineups, one might expect that the TP and TA filler distributions should overlap with each other. However, the TP and TA filler distributions do not necessarily overlap with each other because guilty and innocent suspects differently affect the memory strength of fillers. In general, the memory strength of TP fillers is weaker than that of TA fillers because the guilty suspect makes fillers less appealing than does an innocent suspect[2]—think about the color gray. When you compare it with black, the gray looks brighter than when you compare it with white.

Figure 3 shows the memory-strength distributions of suspects and fillers in empirical eyewitness studies which used either the same fillers (Panel A and C) or different fillers (Panel B and D) in TP and TA lineups. We analyzed Lee, Mansour, and Penrod's (2019) database—for more detailed descriptions of the database, see Supplemental Material 1, assuming that the standard deviation of all memory-strength distributions is equal to 1. In Figure 3, the inverse cumulative distribution function of each identification (ID) rate indicates the distance between the decision criterion and the mean of the memory-strength distribution of the lineup member who is associated with the ID rate. Decision criterion is computed by $-1 \times \{z(\text{guilty suspect ID rate}) + z(\text{innocent suspect ID rate})\}/2$ (Swets, 1973). For example, when a guilty suspect ID rate = .50 (i.e., z(.50) = 0) and the decision criterion = 0.40, the mean of the memory-strength distribution of the guilty suspect is equal to 0.40 (i.e., 0.40 (the criterion) + 0 (the z-value of the guilty suspect ID rate)). Because we are interested in the memory strength of an average filler (rather than the total fillers) in memory-strength distributions, we divided a filler ID rate by the total number of fillers in the lineup and computed the z-value of the average filler ID rate. As shown in Figure 3, filler memory-strength distributions were different between TP and TA lineups, even though the same fillers were used in both TP and TA lineups (Panel A and C). In addition, Panel C and D in Figure 3 also show that innocent suspect and TA filler memory-strength distributions overlapped with each other for studies which assumed a fair TA lineup (i.e., when an innocent suspect and TA fillers have the same memory strength—denoted as average-filler studies).

The different memory-strength distributions between TP and TA fillers for studies which used the same fillers (Panels A and C) implies that memory-strength of lineup members is contextually dependent. When the same fillers are used in TP and TA lineups, the actual similarity of the fillers to the perpetrator is not different between TP and TA lineups; but eyewitnesses perceive the fillers to be less similar to the perpetrator in TP lineups than in TA lineups. More importantly, the contextually dependent memory-strength is also found for guilty and innocent suspects, not only for fillers. The discriminability of a guilty suspect from an innocent suspect varies with context (i.e., compared-to-what), even though the same guilty and innocent suspects were used in the lineups (Bruer, Fitzgerald, Therrien, & Price, 2015; Carlson, Gronlund, & Clark, 2008). We provide an example of the contextually dependent memory-strength of guilty and innocent suspects in Supplement Material 2, by reanalyzing results from Carlson, Gronlund, and Clark's (2008) study.

[2] An alternative explanation about the d' of TP fillers versus TA fillers is present-absent criteria discrepancy theory (Smith, Wells, Lindsay, & Myerson, 2018). According to the theory, the memory strength of TP filler does not differ from that of TA fillers. However, looser criteria in TA lineups (vs. TP lineups) increase TA filler ID rates relative to TP filler ID rates, which leads to the d' of TP fillers versus TA fillers.

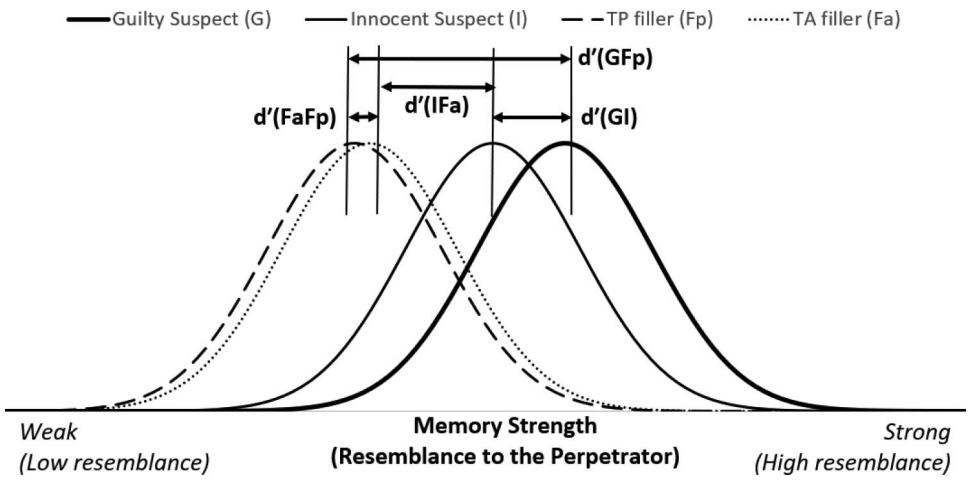


Figure 2. Discriminability measures of eyewitness performance in lineups.

Multiple Discriminability Measures in Previous Eyewitness Research

Compared with d'(GI), the other discriminability measures involving fillers have not been paid much attention in prior studies, although eyewitness research has emphasized the importance of fillers in lineups (Clark, Howell, & Davey, 2008; Wells, 1993). Indeed, some studies propose that the discriminability of a guilty suspect from an innocent suspect is of paramount practical interest, while the discriminability of a guilty/innocent suspect from fillers is unimportant (Wixted & Mickes, 2015a).

Although there have been attempts to incorporate filler IDs into SDT-based discriminability measures, such as *d′-identification* (Horry & Brewer, 2016) and location ROC (LROC; Wixted & Mickes, 2015a), the attempts have limitations. For example, Horry and Brewer (2016) adapted the estimation of d' for n-alternative forced choice (nAFC; Alexander, 2006) to measure the discriminability of a guilty suspect from TP fillers, which is similar to $d'(GF_p)$ in the multi-d' model, and labeled it *d′-identification*. However, because *d′-identification* was originally developed for forced choice tasks, the d' measure is not entirely apt with respect to the discriminability of eyewitnesses who can reject lineups. In addition, *d′-identification* focuses on only the discriminability of a guilty suspect from TP fillers, but not the discriminability of an innocent suspect from TA fillers. In Wixted and Mickes' (2015b) study, LROC measures the discriminability of a guilty suspect from an innocent suspect plus TA fillers. However, the study adapted LROC to justify the use of ROC curves, which focus on the discriminability of a guilty suspect from an innocent suspect, rather to incorporate filler IDs into ROC curves.

Other than developing discriminability measures involving fillers, some studies incorporate filler IDs into SDT measures by including fillers as a parameter in model simulations to estimate underlying d'(GI) from empirical data (Colloff et al., 2016; Palmer & Brewer, 2012; Wixted et al., 2018). Those data simulations estimate not only d'(GI), but also the discriminability of a suspect from fillers. Data simulations that allow multiple discriminability measures are also found in the basic memory literature (e.g., Lampinen, Odegard, Blackshear, & Toglia, 2005). However, because the eyewitness researchers' major interest is to build a best-fitting model to precisely estimate underlying d'(GI) in empirical data, those data simulations focus on the estimation of d'(GI), rather than explaining how the discriminability involving fillers influences the construction of d'(GI).

The neglect of the discriminability measures involving fillers is partly caused by applying the 2 × 2 SDT matrix of yes/no tasks to eyewitness performance in lineups, which is based on a 2 × 3 matrix (Wells et al., 2015; see Table 2). Although filler IDs are technically positive responses, studies wedded to the 2 × 2 framework had to regard them as negative responses like rejections. By combining filler IDs and rejections into negative responses, d'(GI) becomes the only discriminability of interest in prior SDT-based eyewitness studies. Therefore, we propose a new SDT framework for eyewitness performance in lineups incorporating filler IDs; and demonstrate how the multiple discriminability measures involving fillers are related to d'(GI).

Decomposition of d'(GI)

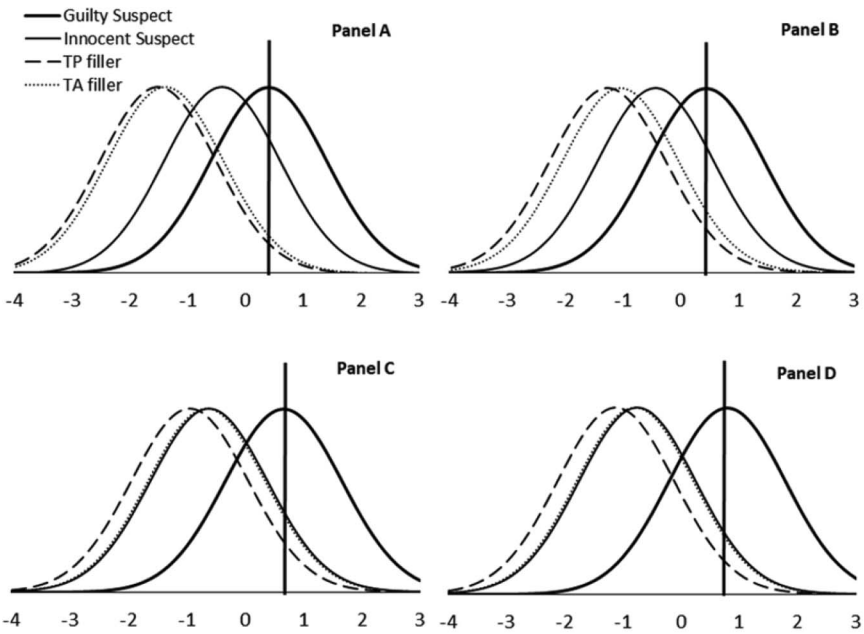
As mentioned earlier, Figure 2 illustrates four different discriminability measures of eyewitnesses viewing lineups. The formulae below express the relationships among the discriminability measures. Given that d' between signal distributions of two stimuli is computed by the z-transformed ID rate of a stimuli minus the z-transformed ID rate of another stimuli (see Footnote 1), $d'(GF_p)$ and $d'(IF_a)$ can be expressed as Formula 1 and 2. Definitions of the acronyms below are provided in Supplemental Material 3.

$$d'(GF_p) = z(G) - z(F_p) \quad \text{(Formula 1)}$$

$$d'(IF_a) = z(I) - z(F_a) \quad \text{(Formula 2)}$$

Considering Formula 1 and 2, $d'(GF_p) - d'(IF_a)$ can be expressed as below.

$$\begin{aligned} d'(GF_p) - d'(IF_a) &= \{z(G) - z(F_p)\} - \{z(I) - z(F_a)\} \\ d'(GF_p) - d'(IF_a) &= z(G) - z(F_p) - z(I) + z(F_a) \\ d'(GF_p) - d'(IF_a) &= \mathbf{z(G)} - \mathbf{z(I)} - z(F_p) + z(F_a) \end{aligned} \quad \text{(Formula 3)}$$



	Innocent-suspect studies				Average-filler studies			
	Same filleres (Panel A)		Different fillers (Panel B)		Same filleres (Panel C)		Different fillers (Panel D)	
	TP	TA	TP	TA	TP	TA	TP	TA
Response Rates								
Suspect ID	0.53	0.23	0.51	0.20	0.46	0.07	0.55	0.07
Filler ID	0.17	0.23	0.24	0.37	0.22	0.40	0.18	0.36
Average Filler ID	0.03	0.04	0.05	0.07	0.04	0.07	0.04	0.07
Rejection	0.30	0.53	0.25	0.43	0.32	0.51	0.27	0.57
Inverse Cumulative Distribution Function								
z(Suspect ID)	0.08	-0.74	0.03	-0.83	-0.11	-1.44	0.11	-1.46
z(Total Fillers ID)	-0.94	-0.72	-0.72	-0.34	-0.76	-0.25	-0.90	-0.36
z(Average Filler ID)	-1.85	-1.72	-1.67	-1.45	-1.73	-1.44	-1.79	-1.46
Criterion	0.33		0.40		0.78		0.67	
Mean of Memory Strength Distribution								
Suspect	0.41	-0.41	0.43	-0.43	0.66	-0.66	0.79	-0.79
Average Filler	-1.52	-1.39	-1.27	-1.05	-0.95	-0.66	-1.12	-0.79
N	11		8		15		2	

Figure 3. Memory-strength distributions of suspects and fillers in prior eyewitness studies. Innocent-suspect studies designated an innocent suspect in TA lineups. Average-filler studies assumed a fair TA lineup and calculated an innocent suspect ID rate by dividing the total filler ID rate in a TA lineup by the total number of persons in the lineup. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

Because z(G) − z(I) is d'(GI), Formula 3 can be simplified as below.

$$d'(GF_p) - d'(IF_a) = d'(GI) - z(F_p) + z(F_a)$$
$$d'(GF_p) - d'(IF_a) = d'(GI) + \mathbf{z(F_a)} - \mathbf{z(F_p)} \quad \text{(Formula 4)}$$

Considering the z-transform equation (see Footnote 1), $z(F_a) - z(F_p)$ can be denoted as $d'(F_aF_p)$. Therefore, with substitutions, Formula 4 can be simplified as below.

$$d'(GF_p) - d'(IF_a) = d'(GI) + d'(F_aF_p)$$
$$d'(GF_p) - d'(IF_a) - d'(F_aF_p) = d'(GI)$$
$$\text{Therefore, } d'(GI) = d'(GF_p) - d'(IF_a) - d'(F_aF_p) \quad \text{(Formula 5)}$$

Formula 5 perfectly fits the results of any empirical eyewitness studies. For example, as shown in Table 3, when comparing the last two columns in the table, the values of d'(GI) and $d'(GF_p) - d'(IF_a) - d'(F_aF_p)$ are identical to each other.

A principle of SDT also predicts Formula 5. According to the additivity of d' in SDT, "the sensitivity statistic d' is a distance measure, and distances along a single dimension add up . . . The sensitivity distance between any stimulus and the endpoint stimulus is a useful measure, *cumulative d'*, that can be computed by adding up adjacent d' values . . . The value of cumulative d' obtained between both endpoint stimuli represents the total sensitivity of the observer to the stimulus set and is called *total d'*. Total

Table 2
A 2 × 3 Matrix of Eyewitness Performance in Prior Eyewitness Studies

	Actual status	
Eyewitness response	Presence of perpetrator	Absence of perpetrator
Suspect ID	True positive	False positive
Filler ID	**False negative**	**True negative**
Rejection	False negative	True negative

d' is the basic measure of observer performance on the entire stimulus ensemble" (Macmillan & Creelman, 2005, p. 114). Given that memory-strength distributions of lineup members are located on the same dimension (i.e., the resemblance to the perpetrator), distances among the distributions can be cumulated. Therefore, in Figure 2, $d'(GF_p)$, which is the total d', equals the sum of $d'(GI)$, $d'(IF_a)$, and $d'(F_aF_p)$. The equation, $d'(GF_p) = d'(GI) + d'(IF_a) + d'(F_aF_p)$, can be transformed to Formula 5, as below.

$$d'(GF_p) = d'(GI) + d'(IF_a) + d'(F_aF_p)$$
$$d'(GF_p) - d'(IF_a) - d'(F_aF_p) = d'(GI)$$
$$\text{Therefore, } d'(GI) = d'(GF_p) - d'(IF_a) - d'(F_aF_p).$$

We suggest that $d'(F_aF_p)$ reflects the differential appeal of fillers between TP and TA lineups. In general, if the same fillers are used in TP and TA lineups, a filler ID rate is higher for TA lineups than for TP lineups because an innocent suspect is less likely to stand out from fillers than the guilty suspect. If an innocent suspect is a clone of the guilty suspect (i.e., if their actual memory-strength distributions have the same μ and σ^2), the value of $d'(F_aF_p)$ is zero. However, such a case is nearly impossible in the real world. Therefore, the value of $d'(F_aF_p)$ is generally positive. Of course, if different fillers are used in TP and TA lineups, the relative appeal of the fillers may vary more dramatically. We denote the index of the differential appeal between TP and TA fillers by $d'(F_aF_p) = z(F_a) - z(F_p)$, rather than $d'(F_pF_a) = z(F_p) - z(F_a)$, to make the index value increase as the differential appeal grows.

In sum, the decomposition of $d'(GI)$ demonstrates that $d'(GI)$ is a function of the discriminability of a suspect from fillers and the differential appeal of fillers between TP and TA lineups.

Multiple Discriminability Measures in Show-Ups and Lineups

Eyewitness identifications in lineups can be regarded as a compound decision task, which is a combination of detection and identification problems (Duncan, 2006). According to Duncan's (2006) signal detection model of compound decision tasks (SDT-CD), eyewitnesses viewing lineups identify a perpetrator (identification decision problem) in the context of uncertainty regarding the presence of the perpetrator (detection decision problem).

In the same vein as SDT-CD, we propose that the decisions eyewitnesses make in lineups can be regarded from two perspective—the discrimination of the presence versus absence of a perpetrator (i.e., detection) and the discrimination of a suspect from fillers (i.e., identification). The second discrimination comprises $d'(GF_p)$, $d'(IF_a)$, and $d'(F_aF_p)$, which consequently yields $d'(GI)$. We argue that the ability relevant to the first discrimination is quantified by the balance of rejection rates between TP and TA lineups. That is, we calculate the discriminability of perpetrator-presence versus perpetrator-absence in lineups as $z(R_a) - z(R_p)$—we denote this parameter by $d'(R_aR_p)$. Penrod (2003) termed the difference in rejection rates a measure of the proportion of "reliable" eyewitnesses—those who could detect the guilty suspect in TP lineups but would reject TA lineups. Horry and Brewer (2016) used a similar index, which was labeled as *d′-detection*, to estimate eyewitnesses' discrimination of the presence versus absence of a perpetrator. Their index was $z(1–R_p) - z(1–R_a)$, which produces the same value of $z(R_a) - z(R_p)$. We can test the validity of $d'(R_aR_p)$ by applying the parameter to eyewitness performance in show-ups and lineups using meta-analytic data from Steblay, Dysart, Fulero, and Lindsay (2003; see Table 4). Of the two types of discriminations, eyewitnesses viewing show-ups are only involved in the discrimination of the presence versus absence of a perpetrator (i.e., whether to reject lineups or not), because show-ups do not include fillers. Therefore, for show-ups, the discriminability of a guilty suspect from an innocent suspect is identical to the discriminability of perpetrator-presence versus perpetrator-absence. Table 4 demonstrates that $d'(GI)$ and $d'(R_aR_p)$ are identical to each other in show-ups.

It is notable that $d'(R_aR_p)$ for show-ups is approximately equal to that for lineups. That is, eyewitnesses' discriminability between target-presence and target-absence was not different for show-ups and lineups. However, compared with show-ups, lineups produced higher discriminability of a guilty suspect from an innocent suspect through the comparisons between a suspect and fillers. As shown in Table 4, $d'(GF_p)$ was higher than $d'(IF_a)$ because of the differential filler siphoning effect (Smith, Wells, Lindsay, & Penrod, 2017; Wells et al., 2015; Wells, Smith, & Smalarz, 2015). The effect refers to the phenomenon that fillers absorb positive IDs more in TA lineups than in TP lineups (Smith et al., 2017; Wells et al., 2015; Wells et al., 2015). That is, as the effect operates more strongly, guilty suspect ID rates increase; TP filler ID rates de-

Table 3
Application of Formula 5 to Results of Prior Eyewitness Studies

Study	Lineup type	G	F_p	R_p	I	F_a	R_a	$d'(GF_p)$	$d'(IF_a)$	$d'(F_aF_p)$	$d'(GI)$	$d'(GF_p) - d'(IF_a) - d'(F_aF_p)$
Carlson and Carlson (2014)	Simultaneous	.32	.44	.24	.06	.64	.30	−.30	−1.87	.50	**1.07**	**1.07**
	Sequential	.25	.55	.20	.07	.69	.24	−.80	−1.99	.37	**.81**	**.81**
Steblay, Dysart, and Wells (2011)	Simultaneous	.52	.25	.24	.28	.26	.46	.72	.06	.03	**.63**	**.63**
	Sequential	.44	.19	.39	.15	.17	.68	.73	−.08	−.08	**.89**	**.89**

Note. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

Table 4
Multiple Discriminability Measures in Show-Ups and Lineups

Discriminability measure	Show-Up		Lineups	
	TP	TA	TP	TA
Suspect ID	.47	.23	.45	.17
Filler ID	NA	NA	.24	.26
Rejection	.53	.77	.31	.57
d'(GI)	.66		.83	
$d'(GF_p)$	NA		.58	
$d'(IF_a)$	NA		−.31	
$d'(F_aF_p)$	NA		.06	
$d'(R_aR_p)$	**.66**		**.67**	

Note. The data of Table 1 in Steblay, Dysart, Fulero, and Lindsay's (2003) meta-analysis was used. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

crease; innocent suspect ID rates decrease; and TA filler ID rates increase, which increases $d'(GF_p) - d'(IF_a)$ and consequently d'(GI). Therefore, the imbalance between $d'(GF_p)$ and $d'(IF_a)$ led to the higher d'(GI) for lineups than for show-ups.

Relationships Among Multiple Discriminability Measures

In this section, we examine relationships among the five d' measures—d'(GI), $d'(GF_p)$, $d'(IF_a)$, $d'(F_aF_p)$, and $d'(R_aR_p)$—with an eyewitness database. We built the database by combining two data sets, Table A.1 in Clark, Howell, and Davey (2008) and a subset of the meta-analysis database of Lee and Penrod (2019)—for more detailed descriptions of the database, see Supplemental Material 4.

First, we looked at correlations and scatterplots of the multiple d' measures (see Figure 4). We separated studies which designated an innocent suspect in TA lineups (innocent-suspect studies) from studies which assumed TA lineups to be perfectly fair by calculating innocent suspect ID rates by dividing the total filler ID rate in a TA lineup by the number of the lineup members (average-filler studies). The two sets of studies may represent distinct eyewitness situations. Innocent-suspect studies may represent situations where an innocent suspect is selected for similarity to the guilty suspect, whereas average-filler studies represent situations in which an innocent suspect arises for reasons other than a general match to the description of a perpetrator (e.g., anonymous tips to police; or a fixed style of committing a crime—modus operandi [MO]). The first situation is more likely to give rise to a biased TA lineup than the latter situation, assuming fillers are selected to match general descriptions of the perpetrator and the appearance of the suspect.

All correlations among the multiple d' measures were comparable for both study types, except the correlation between d'(GI) and $d'(IF_a)$. The correlation between d'(GI) and $d'(IF_a)$ was negative for innocent-suspect studies, $r = -.38$, $p < .001$ while being positive for average-filler studies, $r = .17$, $p < .01$. This pattern arises because innocent suspect ID rates had a negative correlation with d'(GI) for both innocent-suspect and average-filler studies ($r = -.66$ and $-.45$, respectively, ps $< .001$); but the correlation of innocent suspect ID rates and $d'(IF_a)$ was positive for innocent-suspect studies, $r = .75$, $p < .001$—which is not a surprise as many of these lineups were intentionally biased for research purposes; and nonsignificant for average-filler studies, $r = -.10$, $p =$

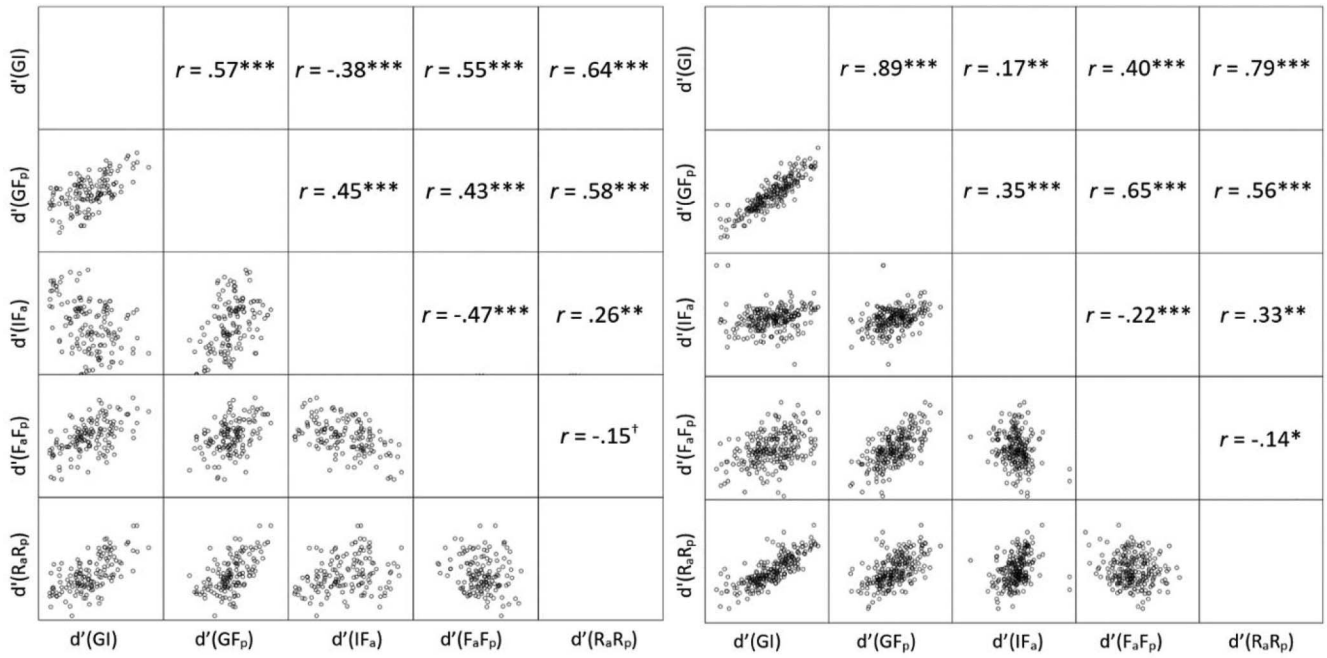


Figure 4. Correlations and scatterplots of multiple d' measures. The left panel was generated from innocent-suspect studies, $n = 146$; the right panel was generated from average-filler studies, $n = 243$. † $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

.14—again, not a surprise as these lineups are analytically presumed to be fair.

The positive correlation between d'(GI) and $d'(F_aF_p)$ in Figure 4 disagrees with Formula 5, which reflects the negative relationship between the two d' measures. We hypothesized that multicollinearity among d' measures might have caused a suppression effect, which changes the correlation sign between d'(GI) and $d'(F_aF_p)$ and conceals their relationship unless other variables are controlled for. As shown in Table 5, when controlling for $d'(GF_p)$ and $d'(IF_a)$, the correlation between d'(GI) and $d'(F_aF_p)$ changed from .55 to a semipartial of −.41 for innocent-suspect studies, and from .40 to a semipartial of −.34 for average-filler studies, which is consistent with Formula 5.

In addition, we conducted regression analyses to investigate the relative influence of $d'(GF_p)$, $d'(IF_a)$, and $d'(F_aF_p)$ on d'(GI). As expected, the three d' predictors explained the variance of d'(GI) completely for both innocent-suspect and average-filler studies; $R^2 = 1.00$ (see Table 5). When comparing βs of the d' predictors, $d'(GF_p)$ and $d'(IF_a)$ affected d'(GI) to a similar degree for innocent-suspect studies, $\beta_{GF_p} = 1.54$ and $\beta_{IF_a} = -1.43$. However, for average-filler studies, when compared with $d'(GF_p)$, $d'(IF_a)$ had less influence on d'(GI), $\beta_{GF_p} = 1.56$ and $\beta_{IF_a} = -0.54$, because the average-filler practice constrains the variability of $d'(IF_a)$, which causes a muted effect of $d'(IF_a)$ on d'(GI).

Application of the Multi-d' Model to Eyewitness Research

Lineup Bias

The resemblance between a suspect and fillers is a crucial factor that influences eyewitness performance. Low resemblance can make the suspect stand out from fillers in a lineup, which increases suspect IDs. Although it would be desirable to construct lineups with high similarity fillers to reduce innocent suspect ID rates, high similarity fillers also induce a decrease in guilty suspect ID rates. Therefore, it is important to find an optimal level of filler similarity to maximize eyewitness performance. Luus and Wells (1991) proposed that the effect of lineup bias on eyewitness performance would follow an inverted U-shape—as lineups become fairer, eyewitness performance increases, but only to a point, beyond which increases in lineup bias lead to a deterioration in eyewitness performance. In the same vein, Fitzgerald, Oriet, and Price (2015) demonstrated that eyewitness performance was poorer in "too fair" lineups (i.e., when fillers extremely resemble the suspect) than in fair lineups. Putting together the notions of Luus and Wells (1991) and Fitzgerald et al. (2015), we can anticipate that the relationship between lineup bias and eyewitness performance would be a skewed inverted-U shape (see Figure 5). In addition, we need to consider a situation where fillers resemble the perpetrator more than does the guilty or innocent suspect, which may not be rare in the real world. Consider, for example, the culprit had a scowled expression during a robbery. Several days later, police officers arrest the robber; compose a lineup with the robber and fillers who match with the victim's descriptions of the robber (e.g., a skinny young White man with blond hair); and present the lineup to the victim. During the lineup procedure, one filler has a scowled expression and others have neutral expressions. In this case, compared with the robber, the scowl filler could be more similar to the victim's memory for the robber's appearance (i.e., a super filler).

Prior studies which have investigated the relationship between lineup bias and eyewitness performance have produced conflicting results (see Table 6). Some studies produced better eyewitness performance (higher d'(GI) and DR) for high similarity lineups, whereas others produced better eyewitness performance for low similarity lineups. In contrast to the conflicting results in d'(GI), $d'(GF_p)$, and $d'(IF_a)$ were generally higher for low similarity lineups than for high similarity lineups. Given that $d'(GF_p)$ and $d'(IF_a)$ reflect the distance between suspect and filler memory-strength distributions (i.e., perceived lineup bias), it does make sense that higher values of $d'(GF_p)$ and $d'(IF_a)$ were associated with low similarity lineups.

Two factors might have caused the unstable pattern of d'(GI) in Table 6. First, the dichotomous comparison (low similarity vs. high similarity) may not reflect the inverted U-shape relationship between lineup bias and eyewitness performance properly because the high- and low-similarity of the studies may rest at different locations on the inverted U-shape curve. Second, eyewitnesses' memory strength of the perpetrator might have an interactive effect with lineup bias on eyewitness performance. Imagine that Eyewitness A and B view the same TP lineup. In this case, the effect of the lineup bias on ID performance would equally operate for the two eyewitnesses. However, when Eyewitness A has stronger memory for the perpetrator, compared with Eyewitness B, Eyewitness A would perceive the TP lineup as more biased, and Eyewitness A would produce better performance (e.g., more likely to identify the guilty suspect). Therefore, we should consider the moderation effect of memory strength of the perpetrator on the relationship between lineup bias and eyewitness performance. For example, compared with other studies, Fitzgerald, Whiting, Therrien, and Price (2014), Study no. 5 in Table 6, produced a much higher $d'(GF_p)$ for both low similarity and high similarity lineups (2.18 for low similarity and 1.68 for high similarity vs. overall averages of 1.75 and 0.54, respectively), which indicates that participants in the study had a strong memory for the perpetrator. Their $d'(IF_a)$ values were comparable with the average $d'(IF_a)$ values (0.37 for low similarity and −0.39 for high similarity vs. the overall 0.54 and −0.66, respectively). In general, $d'(IF_a)$ is not influenced by the memory strength of the perpetrator, because the

Table 5
Regression Analyses of d'(GI) on $d'(GF_p)$, $d'(IF_a)$, and $d'(F_aF_p)$

Predictor	*B*	*SE*	Beta	Zero-order correlation	Semipartial correlation
Innocent-suspect studies					
$d'(GF_p)$	1.00	.00	1.54^{***}	.57	.82
$d'(IF_a)$	−1.00	.00	-1.43^{***}	−.38	−.75
$d'(F_aF_p)$	−1.00	.00	$-.78^{***}$	.55	−.41
R^2			1.00		
Average-filler studies					
$d'(GF_p)$	1.00	.00	1.56^{***}	.89	.88
$d'(IF_a)$	−1.00	.00	$-.54^{***}$	.17	−.39
$d'(F_aF_p)$	−1.00	.00	$-.74^{***}$	.40	−.34
R^2			1.00		

*** $p < .001$.

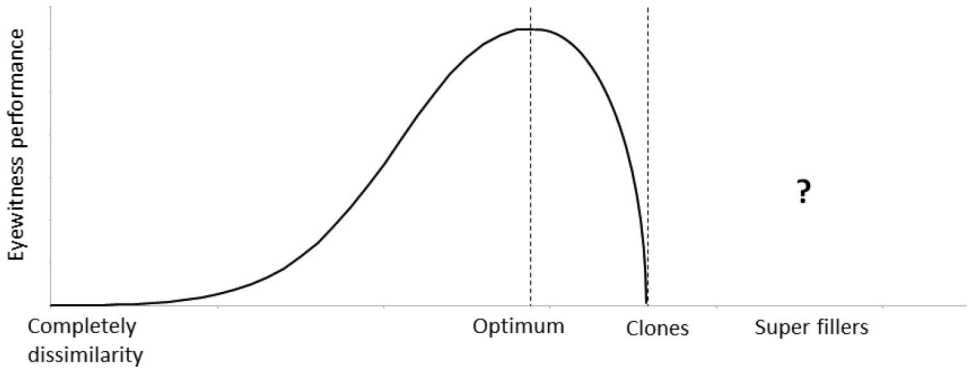


Figure 5. The anticipated relationship between eyewitness performance and filler similarity.

perpetrator is replaced with an innocent suspect in TA lineups (though if an innocent suspect closely resembles the perpetrator, $d'(IF_a)$ could be influenced by the eyewitnesses' memory strength of the perpetrator). That is, in Fitzgerald et al. (2014), the eyewitnesses' strong memory for the perpetrator prevented a substantial decrease in $d'(GF_p)$ when more similar fillers replaced less-similar fillers—but not in $d'(IF_a)$. Therefore, Fitzgerald et al. (2014) produced higher $d'(GI)$ for the high similarity lineup than for the low similarity lineup.

Considering the two factors above, we reclassified the two similarity conditions in Table 6 into four similarity conditions controlling for eyewitnesses' memory strength of the perpetrator. As mentioned earlier, because $d'(IF_a)$ reflects the distance between memory-strength distributions of an innocent suspect and TA fillers who eyewitnesses have not seen before, $d'(IF_a)$ can be regarded as TA lineup bias. However, $d'(GF_p)$ is the product of the memory strength for the perpetrator and TP lineup bias. To control for the memory strength in $d'(GF_p)$, we subtracted $d'(R_aR_p)$ from $d'(GF_p)$ because $d'(R_aR_p)$, the discriminability of perpetrator-presence versus perpetrator-absence, reflects memory strength for the perpetrator. Therefore, we regarded $d'(GF_p) - d'(R_aR_p)$ as TP lineup bias (for more details, see Supplemental Material 5).

In Table 7, we reclassified the two similarity conditions into four similarity conditions based on the mean of $d'(GF_p) - d'(R_aR_p)$ and $d'(IF_a)$—$M \geq 1$; $1 < M \geq 0$; $0 < M \geq -1$; and $M < -1$. We labeled the four conditions as low similarity, moderate similarity, high similarity, and very high similarity conditions, considering that the practical range of the discriminability of

Table 6
Comparisons of Eyewitness Performance Between Low Similarity and High Similarity Lineups

Study no.	G	F_p	R_p	I	F_a	R_a	$d'(GI)$	$d'(GF_p)$	$d'(IF_a)$	$d'(F_aF_p)$	$d'(R_aR_p)$	DR
Low Similarity Lineups												
1	.47	.11	.42	.05	.23	.72	1.57	1.15	−.91	.49	.78	9.40
2	.80	.08	.13	.06	.12	.82	2.35	2.25	−.32	.23	2.06	12.42
3	.71	.06	.24	.64	.12	.24	.19	2.11	1.53	.38	.00	1.11
4	.44	.28	.28	.30	.12	.58	.37	.43	.65	−.59	.78	1.47
5	.76	.07	.17	.28	.17	.54	1.29	2.18	.37	.52	1.05	2.71
6	.62	.07	.31	.47	.14	.39	.38	1.78	1.01	.40	.22	1.32
7	.64	.12	.23	.32	.29	.39	.82	1.53	.11	.60	.45	1.99
8	.65	.02	.33	.40	.02	.59	.65	2.53	1.90	−.02	.66	1.64
Average	.64	.10	.26	.32	.15	.53	.95	1.75	.54	.25	.75	4.01
High Similarity Lineups												
1	.35	.19	.46	.08	.25	.67	1.02	.49	−.73	.20	.54	4.38
2	.53	.18	.29	.05	.36	.59	1.70	.97	−1.27	.54	.78	10.31
3	.31	.22	.47	.16	.51	.33	.50	.28	−1.02	.80	−.36	1.94
4	.16	.61	.23	.24	.38	.38	−.29	−1.27	−.40	−.58	.43	.67
5	.74	.15	.10	.15	.26	.59	1.68	1.68	−.39	.39	1.51	4.93
6	.41	.31	.28	.36	.32	.31	.13	.27	.11	.03	.09	1.14
7	.43	.29	.28	.15	.40	.46	.88	.39	−.79	.29	.48	2.96
8	.54	.07	.39	.10	.31	.59	1.36	1.54	−.77	.95	.51	5.22
Average	.43	.25	.31	.16	.35	.49	.87	.54	−.66	.33	.50	3.94

Note. Results only involving simultaneous lineups and adult participants were included for a precise comparison of d' measures between low similarity and high similarity lineups. Study no. 1: Bergold and Heaton (2018); Study no. 2: Bruer, Fitzgerald, Therrien, and Price (2015); Study no. 3: Carlson, Gronlund, and Clark (2008); Study no. 4: Fitzgerald, Oriet, and Price (2015); Study no. 5: Fitzgerald, Whiting, Therrien, and Price (2014; Exp. 2, adults); Study no. 6: Gronlund, Carlson, Dailey, and Goodsell (2009); Study no. 7: Key, Cash, Neuschatz, Price, Wetmore, and Gronlund (2015, young and middle adults); and Study no. 8: Key, Wetmore, Neuschatz, Gronlund, Cash, and Lane (2017). The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

Table 7
Comparisons of Eyewitness Performance Among the Rearranged Filler Similarity Conditions

Study no.	M of $d'(GF_p) - d'(R_aR_p)$ & $d'(IF_a)$	d'(GI)	$d'(GF_p)$	$d'(IF_a)$	$d'(F_aF_p)$	$d'(R_aR_p)$	DR	$d'(GF_p) - d'(IF_a)$
			Low Similarity Lineups ($M \geq 1$)					
3	1.82	.19	2.11	1.53	.38	.00	1.11	.57
6	1.28	.38	1.78	1.01	.40	.22	1.32	.78
8	1.88	.65	2.53	1.90	−.02	.66	1.64	.63
Average	1.66	.41	2.14	1.48	.25	.29	1.36	.66
			Moderate Similarity Lineups ($1 < M \geq 0$)					
4	.15	.37	.43	.65	−.59	.78	1.47	−.22
5	.75	1.29	2.18	.37	.52	1.05	2.71	1.81
6	.15	.13	.27	.11	.03	.09	1.14	.16
7	.60	.82	1.53	.11	.60	.45	1.99	1.43
8	.13	1.36	1.54	−.77	.95	.51	5.22	2.30
Average	.36	.80	1.19	.09	.30	.58	2.51	1.10
			High Similarity Lineups ($0 < M \geq -1$)					
1	−.27	1.57	1.15	−.91	.49	.78	9.40	2.06
1	−.39	1.02	.49	−.73	.20	.54	4.38	1.22
2	−.06	2.35	2.25	−.32	.23	2.06	12.42	2.57
2	−.55	1.70	.97	−1.27	.54	.78	10.31	2.24
3	−.19	.50	.28	−1.02	.80	−.36	1.94	1.30
5	−.11	1.68	1.68	−.39	.39	1.51	4.93	2.07
7	−.44	.88	.39	−.79	.29	.48	2.96	1.18
Average	−.29	1.39	1.03	−.78	.42	.83	6.62	1.81
			Very High Similarity Lineups ($M < -1$)					
4	−1.05	−.29	−1.27	−.40	−.58	.43	.67	−.87

a suspect from fillers (i.e., either $d'(GF_p)$ or $d'(IF_a)$) would be roughly between −1.91 and 4.65 (for more detailed explanations, see Supplemental Material 6).

As expected, the filler similarity effect on eyewitness performance followed a skewed inverted-U shape relationship. In Table 7, d'(GI) increased until reaching the high similarity lineup (0.41, 0.80, and 1.39); but decreased in the very high similarity lineup (−0.29). The trend of $d'(R_aR_p)$ and DR also followed a skewed inverted-U shape. $d'(R_aR_p)$ and DR reached the maximum value in high similarity lineups.

The effect of lineup bias on d'(GI) could be largely accounted for by $d'(GF_p) - d'(IF_a)$. Given the multi-d' model, $d'(GI) = d'(GF_p) - d'(IF_a) - d'(F_aF_p)$, a larger imbalance between $d'(GF_p)$ and $d'(IF_a)$ should be associated with higher d'(GI). In Table 7, $d'(GF_p) - d'(IF_a)$ also had the skewed inverted U-shape relationship with lineup bias—the value increased until high similarity lineup (0.66, 1.10, and 1.81); but decreased in very high similarity lineup (−0.87). As mentioned earlier, we suggest that $d'(GF_p) - d'(IF_a)$ is an index of the differential filler siphoning effect. As the effect operates more strongly, guilty suspect ID rates increase; TP filler ID rates decrease; innocent suspect ID rates decrease; and TA filler ID rates increase, which increases $d'(GF_p) - d'(IF_a)$ and consequently d'(GI). It is notable that $d'(F_aF_p)$, which reflects stronger appeals of TA fillers than TP fillers, also followed the skewed inverted-U shape because $d'(F_aF_p)$ is closely related to the differential filler siphoning effect—as the effect operates more strongly, $d'(F_aF_p)$ increases.

Our analyses in Table 7 indicate there is an optimal level of lineup bias, which maximizes eyewitness performance. However, it is still unclear whether lineup bias has the anticipated relationship with eyewitness performance in Figure 5, because the included studies had a narrow range of lineup bias (the range of $d'(GF_p)$ and $d'(IF_a)$ was from −1.27 to 2.53). Therefore, we conducted computational simulations to investigate the relationship between lineup bias and eyewitness performance with the full range of the d' measures.

Figure 6 illustrates the *actual* memory strength of suspects and a filler ($M_{guilty} = 1$, $M_{innocent} = 0$, $M_{filler} = -1$, all SDs = 1). We can compute the probability of positive IDs for each of the distributions (i.e., shaded areas) because the memory strength distributions follow a normal distribution function. For example, when criterion = 0, the guilty suspect ID probability = .84; the innocent suspect ID probability = .50; and the filler ID probability = .16 × 5 = .80 (because the filler distribution reflects only one filler's memory strength, we multiply the filler ID probability by the number of fillers assuming a six-person lineup). We propose that the proportion of the filler ID probability in the total positive ID probability reflects the filler siphoning effect. That is, in Figure 6, fillers absorb 49% (=0.80/(0.80 + 0.84)) of the positive IDs in TP lineups, and 61% (=0.80/[0.80 + 0.50]) of the positive IDs in TA lineups. More importantly, the difference in the proportions of filler ID probability in the total positive ID probability between TP and TA lineups would reflect the *differential* filler siphoning effect. In Figure 6, TA fillers (vs. TP fillers) absorb 12% (=61% − 49%) more positive IDs.

Following this computation method, we can quantify the differential filler siphoning effect, varying the memory strength of fillers. As shown in Figure 7, the effect grows stronger as fillers

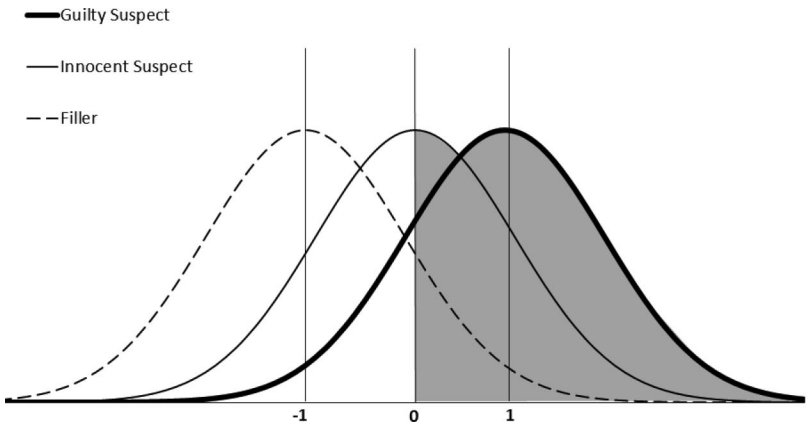


Figure 6. Actual memory-strength distributions of suspects and a filler.

more closely resemble the suspect. However, there is a tipping point beyond which increases in fillers' memory strength decreases the effect. In this simulation, when $M_{filler} = -1.10$, the difference in the proportion of the filler ID probability in the total positive ID probability between TP and TA lineups reached the maximum value (0.129). Beyond the clone similarity level (i.e., $M_{filler} = M_{guilty} = 1$), the effect slowly decreases. The trend of the differential filler siphoning effect in Figure 7 is similar to the anticipated relationship between lineup bias and eyewitness performance in Figure 5, except that the effect at the clone similarity level is not as weak as that at the completely dissimilar level. We argue that the relationship between lineup bias and eyewitness performance would follow the trend of the differential filler siphoning effect in Figure 7, because eyewitness performance (d'(GI)) is computed with $d'(GF_p) - d'(IF_a)$ and $d'(F_aF_p)$, which are closely related to the effect. Therefore, the differential filler siphoning effect is the mechanism that makes d'(GI) vary with filler similarity when the actual memory-strength of guilty and innocent suspects does not change.

Figure 8 and 9 illustrate the trend of the differential filler siphoning effect, varying eyewitnesses' criteria and the resemblance of guilty and innocent suspects. As shown in Figure 8, when eyewitnesses use stringent criteria (vs. loose criteria), the effect is stronger and more sensitive to filler similarity. When an innocent suspect does not bear a resemblance to the perpetrator, the effect operates more strongly (Panel B in Figure 9). However, when an innocent suspect is a clone of the perpetrator (Panel A in Figure 9), the effect does not occur.

Filler Selection Methods

Prior eyewitness studies have investigated which filler-selection method (match-to-suspect vs. match-to-description) produces better eyewitness performance (Juslin, Olsson, & Winman, 1996;

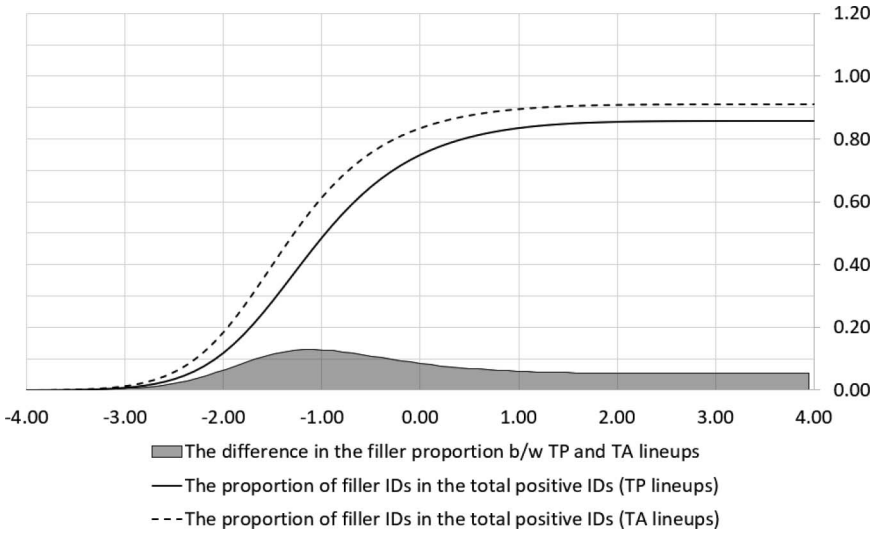


Figure 7. The trends of the differential filler siphoning effect in lineups ($M_{guilty} = 1$, $M_{innocent} = 0$, all SDs = 1, and criterion = 0). The x-axis represents the memory strength of fillers. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

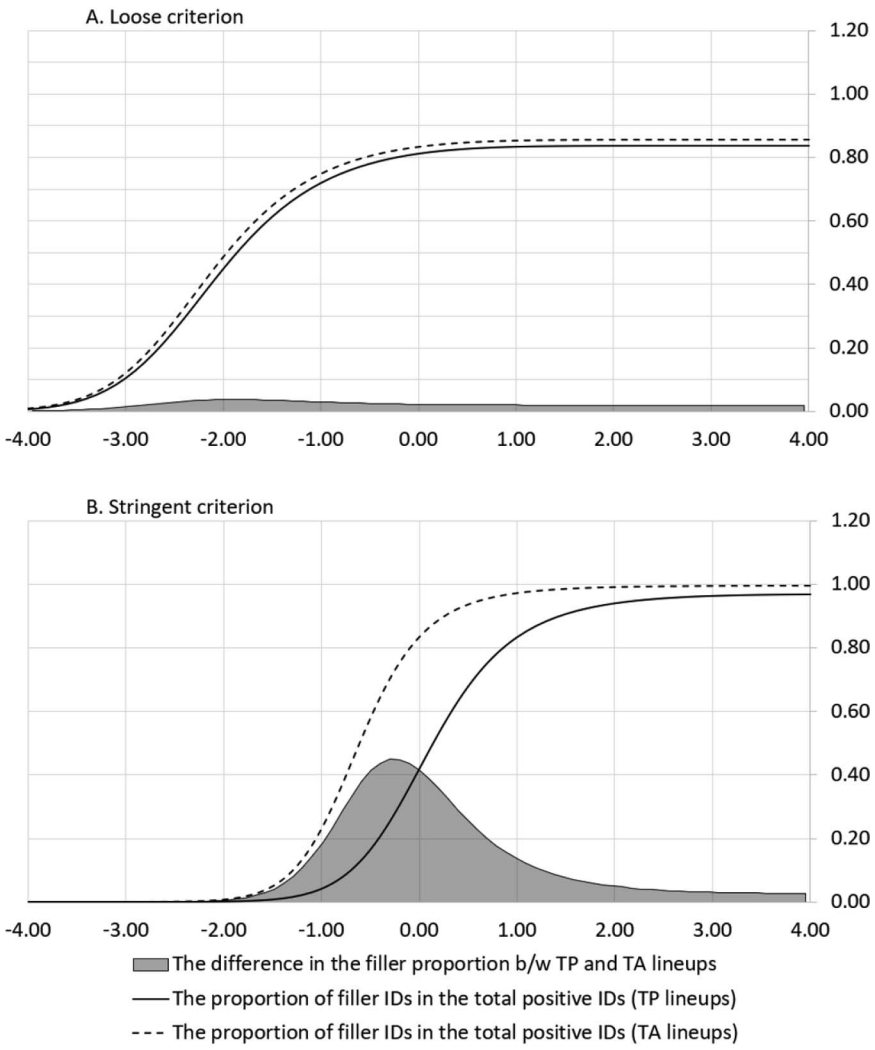


Figure 8. Eyewitness criteria and the trends of the differential filler siphoning effect in lineups ($M_{\text{guilty}} = 1$, $M_{\text{innocent}} = 0$, all *SD*s = 1). Panel A was drawn when criterion = −1; Panel B was drawn when criterion = 2. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

Tunnicliff & Clark, 2000; Wells, Rydell, & Seelau, 1993). The match-to-suspect method selects fillers based on their similarity to the suspect, whereas the match-to-description method selects fillers based on their similarity to the eyewitnesses' descriptions of the perpetrator (Luus & Wells, 1991). As shown in Table 8, Tunnicliff and Clark (2000) found higher d'(GI) for the match-to-suspect condition, whereas d'(GI) in Juslin, Olsson, and Winman (1996) and Wells, Rydell, and Seelau (1993) was higher for the match-to-description condition.

To explain the conflicting results, we looked at the multiple discriminability measures. The match-to-description method generally yielded more biased TP and TA lineups (which increase $d'(GF_p)$ and $d'(IF_a)$), than did the match-to-suspect method. However, the advantage of $d'(GF_p)$ for the match-to-description method was greater in Juslin et al.'s (1996) and Wells et al.'s (1993) studies (vs. Tunnicliff & Clark's, 2000 studies), whereas the advantage of $d'(IF_a)$ for the match-to-description method was not considerably different across the four studies. Therefore, d'(GI) was higher for the match-to-description lineups in Juslin et al.'s (1996) and Wells et al.'s (1993) studies, but not in Tunnicliff and Clark's (2000) studies.

Based on these results, we hypothesize that, compared with Juslin et al.'s (1996) and Wells et al.'s (1993) studies, Tunnicliff and Clark's (2000) studies may have used higher-quality descriptions about the perpetrator's appearance. With better descriptions, the advantage of $d'(GF_p)$ for the match-to-description method might have disappeared in Tunnicliff and Clark's (2000) studies. Indeed, to compose match-to-description lineups, eyewitnesses in Tunnicliff and Clark's (2000) studies were asked to write down more elaborated descriptions of the perpetrator after they wrote down their initial descriptions. In contrast, the perpetrator-description in Juslin et al.'s (1996) study included few characteristics—age, sex, race, hair type, and body weight. Wells et al.'s (1993) study, which yielded the highest $d'(GF_p)$ for the match-to-description method, disregarded photos that violated the description of the perpetrator from a photo pool; and then selected fillers

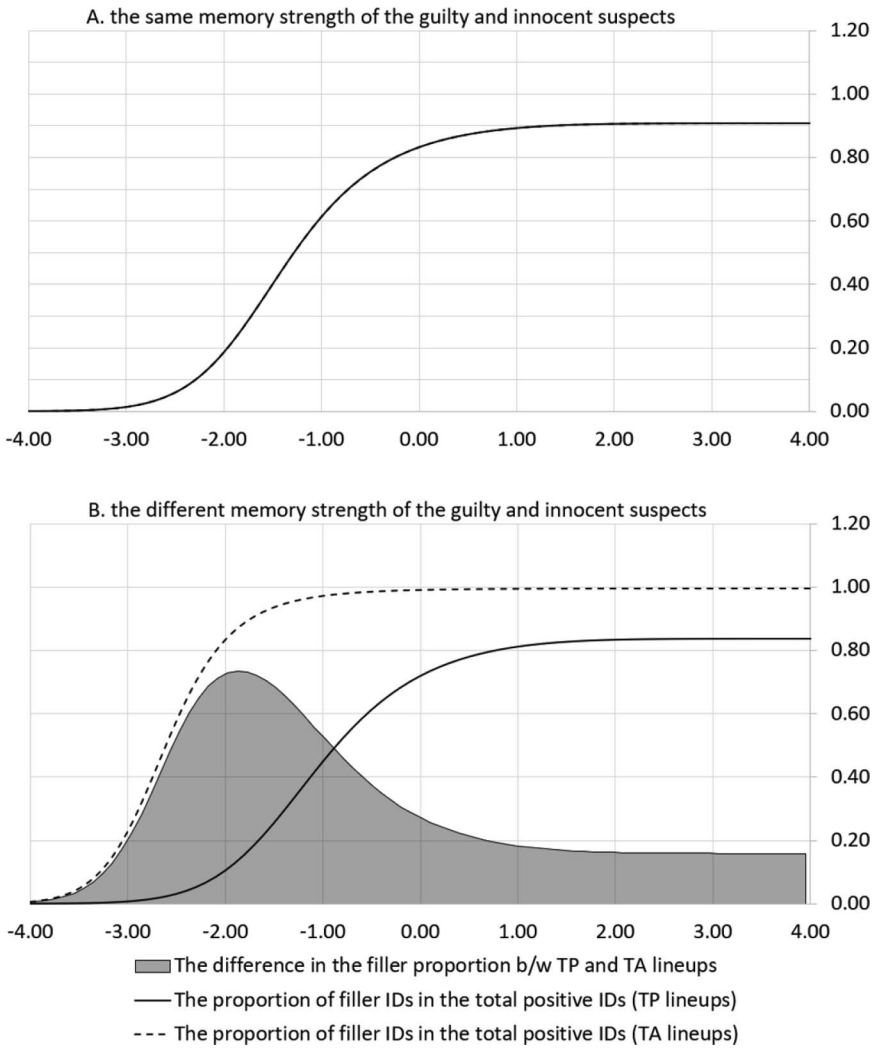


Figure 9. The resemblance between guilty and innocent suspects and the trends of the differential filler siphoning effect in lineups (criterion = 0). Panel A was drawn when $M_{\text{guilty}} = 0$, $M_{\text{innocent}} = 0$; Panel B was drawn when $M_{\text{guilty}} = 2$, $M_{\text{innocent}} = -2$. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

who least resembled the perpetrator from the remaining photos to compose the match-to-description lineup.

Therefore, the multi-d' model indicates that the effect of filler-selection methods on eyewitness performance will vary with the quality of the descriptions. Results from the four studies also raise the question of what the optimal level of resemblance between a suspect and fillers is to maximize d'(GI). We expect our multiple discriminability measures could be useful metrics to estimate the resemblance among a suspect and fillers, to identify optimal levels and to specify the effects of more and less-optimal similarities.

Eyewitness Confidence

SDT explains that discriminability is independent of criterion (Swets, 1973). That is, unlike Bayesian measures (e.g., DR), SDT-based discriminability measures (e.g., d') do not vary with criteria. We can illustrate and qualify these points using data from Evelo, Lee, Modjadidi, and Penrod (2018)—for more detailed descriptions of the database, see Supplemental Material 7. In Table 9, we collapse across the independent variables (except the target presence in lineups) and present the eyewitness performance measures, which were produced from about 16,000 identification tasks (2,000 participants × 8 lineups).

As shown in Table 9 (where judgments cumulate starting with the percentage of participants making judgments at 100% confidence, then the percent at 90% or higher, 80% and higher and so on), the criterion parameter (c) varied with eyewitnesses' confidence levels—as eyewitnesses were less confident, the value of c decreased as expected. Compared with DR, d'(GI) and $d'(R_aR_p)$ were, as expected, relatively stable across confidence levels. However, the three other discriminability measures were influenced by eyewitnesses' criteria and produced countervailing effects. Eyewitnesses' looser criteria were associated with lower discrim-

Table 8
Eyewitness Performance in Match-to-Suspect and Match-to-Description Lineups

Study	Method	G	F_p	R_p	I	F_a	R_a	d'(GI)	$d'(GF_p)$	$d'(IF_a)$	$d'(F_aF_p)$	$d'(R_aR_p)$
Tunnicliff and Clark (2000) Exp. 1	Suspect-Match	.53	.25	.22	.03	.31	.66	1.94	.75	−1.38	.19	1.18
	Description-Match	.53	.16	.31	.13	.34	.53	1.23	1.09	−.75	.61	.57
Tunnicliff and Clark (2000) Exp. 2	Suspect-Match	.33	.27	.40	.08	.19	.73	.95	.18	−.50	−.28	.87
	Description-Match	.31	.25	.44	.19	.35	.46	.40	.19	−.51	.30	.05
Juslin, Olsson, and Winman (1996)	Suspect-Match	.44	.20	.35	.09	.17	.73	1.19	.69	−.39	−.11	1.00
	Description-Match	.52	.11	.38	.09	.12	.78	1.39	1.28	−.17	.05	1.08
Wells, Rydell, and Seelau (1993)	Suspect-Match	.21	.43	.36	.12	.48	.40	.37	−.63	−1.12	.13	.11
	Description-Match	.67	.07	.26	.12	.31	.57	1.61	1.92	−.68	.98	.82

Note. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

inability of a guilty/innocent suspect from fillers. $d'(GF_p)$ and $d'(IF_a)$ decreased as eyewitnesses were less confident. Higher $d'(F_aF_p)$ values were also associated with lower confidence levels because the stronger appeal of fillers relative to suspects in TA lineups than in TP lineups is more likely to occur for eyewitnesses who make their decisions by guessing/less reliable memories (i.e., eyewitnesses at low confidence levels).

Figure 10 is an area graph which cumulates values of $d'(GF_p)$, $-d'(IF_a)$, and $-d'(F_aF_p)$ at each level of confidence in Table 9. The area graph illustrates how the different components make varying contributions to d'(GI) over different confidence levels—as eyewitnesses use a looser criterion, the contribution of $d'(GF_p)$ decreases; and the contribution of $d'(IF_a)$ increases. d'(GI) at the level of 100% confidence was 1.09, which comprised $d'(GF_p)$ of 0.77; $d'(IF_a)$ of −0.36; and $d'(F_aF_p)$ of 0.03 (i.e., $1.09 = 0.77 - (-0.36) - 0.03$). That is, d'(GI) at the highest level of confidence was largely accounted for by high $d'(GF_p)$. Although d'(GI) at the level of 0% confidence (0.97) was comparable with that at the level of 100% confidence (1.09), eyewitnesses with 0% confidence achieved the d'(GI) mostly by low $d'(IF_a)$; d'(GI) at the level of 0% confidence comprised $d'(GF_p)$ of −0.10; $d'(IF_a)$ of −1.29; and $d'(F_aF_p)$ of 0.22 (i.e., $0.97 = -0.10 - (-1.29) - 0.22$).

In sum, eyewitnesses show stable d'(GI) across confidence levels. However, the quality of d'(GI) is varied by confidence levels. Eyewitnesses with high confidence achieve good d'(GI) by their good discriminability of a guilty suspect from fillers, whereas those with low confidence show the same d'(GI) by their poor discriminability of an innocent suspect from fillers (their "guessing" protects the innocent suspect and thereby increases d'(GI)), rather than a good discriminability of a guilty suspect from fillers.

Lineup Presentation Mode

Eyewitness studies have traditionally used Bayesian measures, such as DR or conditional probability, to assess eyewitness accuracy. Prior studies using Bayesian probabilities supported the superiority of sequential lineups to simultaneous lineups by showing higher DRs for sequential lineups (e.g., Cutler & Penrod, 1989; Lindsay et al., 1991; Steblay, Dysart, & Wells, 2011). The superiority of sequential lineups has been challenged recently by studies based on a different form of a diagnostic measure—receiver operating characteristic (ROC) analysis (e.g., Gronlund, Carlson, Dailey, & Goodsell, 2009; Mickes, Flowe, & Wixted, 2012). Skeptics of the superiority of sequential lineups assert that sequential lineups (vs. simultaneous lineups) produce higher DR because eyewitnesses in sequential lineups are more likely to make their decisions at stringent criteria. Therefore, researchers began to compare eyewitness diagnosticity in both lineups across multiple criteria using ROC analysis.

In this section, we propose that the multi-d' model is a useful tool to compare the underlying process of eyewitness performance between simultaneous and sequential lineups. Because both Bayesian measures and ROC analysis focus on guilty and innocent suspect IDs only, they have rarely paid attention to the underlying process of eyewitness performance involving fillers. Table 10

Table 9
Eyewitness Performance and Criteria Across Confidence Levels (Frequencies in Parentheses)

Confidence levels	G	F_p	R_p	I	F_a	R_a	d'(GI)	$d'(GF_p)$	$d'(IF_a)$	$d'(F_aF_p)$	$d'(R_aR_p)$	DR	c
100%	.03 (243)	.004 (33)	.02 (123)	.002 (12)	.005 (36)	.04 (304)	1.09	.77	−.36	.03	.39	20.25	2.42
90%	.05 (415)	.01 (66)	.03 (235)	.003 (25)	.01 (82)	.06 (511)	1.11	.77	−.42	.08	.37	16.60	2.18
80%	.09 (710)	.02 (186)	.05 (398)	.01 (77)	.03 (237)	.10 (826)	.99	.64	−.45	.10	.38	9.22	1.84
70%	.13 (1,046)	.05 (416)	.08 (613)	.02 (128)	.07 (527)	.15 (1,204)	1.02	.50	−.64	.12	.39	8.17	1.63
60%	.17 (1,391)	.10 (768)	.10 (821)	.02 (191)	.12 (933)	.20 (1,558)	1.04	.37	−.79	.11	.41	7.28	1.46
50%	.22 (1,769)	.16 (1,270)	.15 (1,201)	.04 (296)	.19 (1,551)	.26 (2,082)	1.02	.23	−.92	.14	.39	5.98	1.28
40%	.25 (1,998)	.21 (1,652)	.18 (1,423)	.05 (384)	.26 (2,051)	.30 (2,373)	.99	.14	−1.01	.16	.39	5.20	1.17
30%	.28 (2,256)	.26 (2,058)	.21 (1,652)	.06 (456)	.33 (2,601)	.34 (2,718)	1.00	.08	−1.13	.20	.41	4.95	1.08
20%	.31 (2,444)	.31 (2,439)	.23 (1,860)	.07 (542)	.38 (3,046)	.37 (2,984)	.98	.00	−1.19	.21	.41	4.51	1.00
10%	.32 (2,567)	.34 (2,713)	.26 (2,043)	.07 (596)	.42 (3,372)	.41 (3,235)	.98	−.05	−1.25	.22	.42	4.31	.95
0%	.33 (2,622)	.37 (2,922)	.31 (2,437)	.08 (632)	.45 (3,613)	.47 (3,735)	.97	−.10	−1.29	.22	.43	4.15	.93

Note. The excel spreadsheet containing the computational formulas is available on OSF website (http://bit.ly/Multi_d).

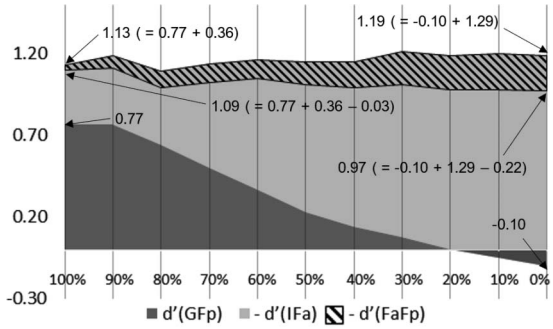


Figure 10. The contribution of each d' measure to d'(GI).

includes eyewitnesses' response rates and multiple discriminability measures for simultaneous and sequential lineups, which were retrieved from Steblay, Dysart, and Wells' (2011) meta-analysis dataset and Lee and Penrod's (2019) meta-analysis dataset (for more detailed descriptions of the database, see Supplemental Material 8). We separated innocent-suspect studies from average-filler studies to examine differences between both lineup presentation modes.

As shown in Table 10, for innocent-suspect studies, eyewitness performance in sequential lineups was better than that in simultaneous lineups in terms of d'(GI), $d'(R_aR_p)$, and DR, which are indicators of overall eyewitness performance. Eyewitnesses in simultaneous lineups (vs. those in sequential lineups) were slightly better at discriminating the guilty suspect from fillers ($d'(GF_p)$ = 0.95 vs. 0.91); but were substantially more likely to misidentify an innocent suspect from fillers ($d'(IF_a) = -0.04$ vs. -0.24), which consequently yielded lower d'(GI) in simultaneous lineups. Given that lineups in innocent-suspect studies were more biased than those in average-filler studies (for more detailed explanations, see Supplemental Material 9), the superiority of sequential lineups to simultaneous lineups for innocent-suspect studies might arise because sequential lineups produce better eyewitness performance when lineups are biased. Compared with simultaneous lineups, sequential lineups are more likely to reduce lineup bias (Lindsay et al., 1991) and consequently the discriminability of a suspect from fillers—indeed, in Table 10, $d'(GF_p)$ and $d'(IF_a)$ for innocent-suspect studies were lower in sequential lineups (0.91 and -0.24, respectively) than simultaneous lineups (0.95 and -0.04, respectively). It is notable that sequential lineups reduce the discriminability of a suspect from fillers more in TA lineups compared with TP lineups, because the presence of the perpetrator in TP lineups protects against the decrease in $d'(GF_p)$. As expected, when comparing sequential lineups to simultaneous lineups, $d'(GF_p)$ decreased by 0.04 ($=0.95 - 0.91$) while $d'(IF_a)$ decreased by 0.20 ($=-0.04 - (-0.24)$). Therefore, in innocent-suspect studies, sequential lineups produced superior d'(GI) compared to simultaneous lineups.

For average-filler studies which used fairer lineups than did innocent-suspect studies (see Supplemental Material 9), d'(GI) and $d'(R_aR_p)$ were higher for simultaneous lineups than for sequential lineups. Although $d'(GF_p)$ was higher in simultaneous lineups than in sequential lineups for both innocent-suspect and average-filler studies, the difference in $d'(GF_p)$ between the two lineup modes was larger for average-filler studies (0.95 vs. 0.91 for innocent-suspect studies; 0.89 vs. 0.63 for average-filler studies). That is, the superiority of simultaneous lineups in $d'(GF_p)$ increases for fairer lineups. $d'(IF_a)$ for average-filler studies was lower in simultaneous lineups than in sequential lineups because the average-filler practice makes sequential lineups produce higher $d'(IF_a)$—the average-filler practice makes lineups with higher rejection rates produce higher $d'(IF_a)$, regardless of the lineup bias. As mentioned in Supplemental Material 6, when using the average-filler practice (i.e., dividing the total filler ID rate by the number of members in the lineup), $d'(IF_a)$ in a lineup gradually increases as the rejection rate increases. For example, when a rejection rate is zero, $d'(IF_a)$ is -1.91; when a rejection rate is .10, $d'(IF_a)$ is -1.71; when a rejection rate is .20, $d'(IF_a)$ is -1.57. Indeed, when applying the average-filler practice to innocent-suspect studies (see debiased studies in Table 10 where it is analytically assumed TA arrays are unbiased), $d'(IF_a)$ was higher for sequential lineups (-1.12) than for simultaneous lineups (-1.25), although simultaneous TA line-

Table 10
Eyewitness Performance in Simultaneous and Sequential Lineups (SDs in Parentheses)

Eyewitness performance	Innocent-suspect studies				Average-filler studies				Debiased studies			
	Simultaneous		Sequential		Simultaneous		Sequential		Simultaneous		Sequential	
	TP	TA	TP	TA	TP	TA	TP	TA	TP	TA	TP	TA
Suspect ID	.54 (.14)	.28 (.12)	.47 (.16)	.19 (.13)	.55 (.17)	.08 (.02)	.40 (.16)	.05 (.02)	.54 (.14)	.09 (.02)	.47 (.16)	.07 (.04)
Filler ID	.22 (.14)	.29 (.16)	.21 (.15)	.25 (.18)	.23 (.09)	.41 (.09)	.20 (.13)	.27 (.10)	.22 (.14)	.47 (.12)	.21 (.15)	.37 (.18)
Rejection	.24 (.08)	.44 (.15)	.32 (.15)	.56 (.22)	.22 (.13)	.50 (.11)	.40 (.16)	.67 (.12)	.24 (.08)	.44 (.15)	.32 (.15)	.56 (.22)
d'(GI)	.76 (.64)		.95 (.74)		1.51 (.51)		1.36 (.52)		1.43 (.48)		1.45 (.64)	
$d'(GF_p)$	.95 (.85)		.91 (.95)		.89 (.69)		.63 (.74)		.95 (.85)		.91 (.95)	
$d'(IF_a)$	−.04 (.81)		−.24 (.80)		−1.18 (.11)		−1.00 (.13)		−1.25 (.17)		−1.12 (.24)	
$d'(F_aF_p)$	.24 (.43)		.20 (.42)		.55 (.27)		.26 (.28)		.77 (.38)		.57 (.39)	
$d'(R_aR_p)$	.55 (.42)		.71 (.64)		.82 (.51)		.78 (.49)		.55 (.42)		.71 (.64)	
$d'(GF_p) - d'(R_aR_p)$	.40 (.70)		.20 (.65)		.07 (.49)		−.15 (.63)		.40 (.70)		.20 (.65)	
DR	3.79 (7.46)		5.06 (5.71)		7.11 (3.42)		8.94 (5.22)		6.55 (3.93)		10.79 (13.89)	
c	.28 (.28)		.53 (.32)		.64 (.23)		.96 (.22)		.62 (.16)		.78 (.20)	
N	21		21		20		20		21		21	

Note. Eyewitness performance in debiased studies were computed by applying the average-filler practice to innocent suspect and TA filler ID rates in innocent-suspect studies. For example, TA suspect ID rates in simultaneous lineups for debiased studies were computed by (.28 + .29)/6 = .09.

ups were actually more biased than sequential TA lineups for innocent-suspect studies. Because sequential lineups, compared to simultaneous lineups, induce more stringent criteria (see *c* in Table 10), sequential lineups yield higher rejection rates and consequently higher $d'(IF_a)$ when using the average-filler practice. Therefore, as expected, $d'(IF_a)$ was higher for sequential lineups (−1.00) than for simultaneous lineups (−1.18).

In sum, as shown in Figure 11, for innocent-suspect studies, simultaneous lineups (vs. sequential lineups) produced a slightly higher $d'(GF_p)$ and substantially higher $d'(IF_a)$, which yielded lower $d'(GI)$ in simultaneous lineups. For average-filler studies, the advantage of $d'(GF_p)$ in simultaneous lineups (vs. sequential lineups) increased; and the average-filler practice led to lower $d'(IF_a)$ in simultaneous lineups, which consequently yielded higher $d'(GI)$ in simultaneous lineups.

Conclusions

Eyewitness studies have leveraged SDT to investigate eyewitness performance. The traditional SDT-based measures in yes/no tasks properly captures eyewitness performance in show-ups, but not in lineups, because the application of the measures to eyewitness identifications was based on the 2 × 2 matrix, which led to the neglect of the role of fillers in lineups. Although there have been attempts to incorporate filler IDs into SDT-based measures, most prior SDT-based eyewitness studies have focused on $d'(GI)$, ignoring other discriminability measures involving fillers; and have limited our understanding of the role of fillers in eyewitness performance.

In the current research, we introduced a SDT-based framework for eyewitness performance in lineups—the multi-d' model. The model demonstrates that $d'(GI)$ is a function of discriminability involving fillers. Furthermore, eyewitnesses' discriminability in lineups can be assessed at two levels—detection and identification levels. At the detection level, an eyewitness discriminates the presence versus absence of the perpetrator in a lineup (i.e., $d'(R_aR_p)$), and makes a decision on whether to reject the lineup or not. At the identification level, the eyewitness is comparing lineup members to choose a person who is most likely to be the perpetrator. At the identification level, discriminability measures involving fillers (i.e., $d'(GF_p)$, $d'(IF_a)$, and $d'(F_aF_p)$) operate and yield the most commonly reported parameter, $d'(GI)$.

Show-Ups and Lineups in the Multi-d' Model

We have demonstrated that the multi-d' model can be applied to issues in eyewitness research and provides useful parameters to investigate eyewitness performance. For example, the multi-d' model explains how different eyewitness performance is in show-ups and lineups. According to the multi-d' model, eyewitnesses' discriminability of the presence versus absence of a perpetrator is not different between show-ups and lineups. However, eyewitnesses in lineups are better at discriminating a guilty suspect from an innocent suspect than those in show-ups, through the differential filler siphoning effect. These findings are consistent with prior studies which suggested the superiority of lineups over show-ups, on the basis of the differential filler siphoning effect. The origin of the superiority of lineups goes back to the idea that eyewitnesses rely on relative judgments to make their decisions in lineups (Wells, 1984). An eyewitness in a show-up has to decide whether to identify the suspect as the perpetrator by comparing the suspect's appearance to the eyewitness' memory for the perpetrator (i.e., absolute judgment), whereas an eyewitness in a lineup compares lineup members with each other and identifies the lineup member who most resembles the eyewitness' memory for the perpetrator (i.e., relative judgment). Because of the relative judgment in lineups, positive identifications in lineups are distributed to fillers and the suspect whereas all positive identifications in show-ups load up on the suspect (Wells, 2001). Given that this filler siphoning effect in lineups grows stronger in TA lineups than TP lineups (i.e., the differential filler siphoning effect), lineups are more likely to reduces the risk of innocent suspect IDs than show-ups (Wells et al., 2015; Wells et al., 2015).

Lineup Bias in the Multi-d' Model

The multi-d' model explains how lineup bias affects eyewitness performance. According to the model, the effect of lineup bias on eyewitness performance follows a skewed inverted U-shape. Eyewitness performance, reflected in $d'(GI)$, $d'(R_aR_p)$, and DR, increases gradually until high similarity lineups are reached; but performance decreases in very high similarity lineups. Results from prior studies and computational simulations demonstrated that the trend of d' measures which are relevant to the differential filler siphoning effect (i.e., $d'(GF_p) - d'(IF_a)$, and $d'(F_aF_p)$) has

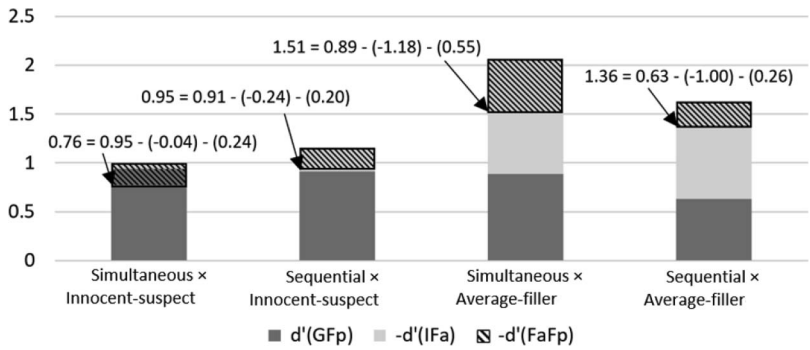


Figure 11. The contribution of each d' measure to $d'(GI)$ in Lineup Mode (simultaneous lineups vs. sequential lineups) × Study Type (innocent-suspect studies vs. average-filler studies). Formula 5, $d'(GI) = d'(GF_p) - d'(IF_a) - d'(F_aF_p)$, was applied to each condition.

the same skewed inverted-U shape with eyewitness performance. These results suggest that the effect of lineup bias on eyewitness performance can be accounted for by the differential filler siphoning effect. This conclusion is consistent with findings from the prior studies on the effect (Smith et al., 2017; Wells et al., 2015; Wells et al., 2015). Those prior studies have suggested that the differential filler siphoning effect grows stronger as lineups become fairer, because good fillers (vs. poor fillers) are more likely to absorb false-positive errors and reduces innocent suspect ID rates more effectively.

Filler Selection Methods in the Multi-d' Model

The multi-d' model explains that the effect of filler-selection methods on eyewitness performance will vary with the quality of eyewitnesses' descriptions of the perpetrator. Compared with the match-to-suspect method, the match-to-description method generally produces more biased TP and TA lineups. However, when lineups are composed by very precise descriptions of the perpetrator, the TP lineups are no longer more biased than TP lineups composed by the match-to-suspect method; but the match-to-description TA lineups are still more biased than TA lineups composed by the match-to-suspect method. Therefore, d'(GI) is lower in lineups composed by precise descriptions of the perpetrator than in lineups composed by the match-to-suspect method. These findings raise the question of how detailed descriptions of the perpetrator in real cases are. Depending on how detailed they are, we might prefer one versus the other method of filler selection.

Eyewitness Confidence in the Multi-d' Model

The multi-d' model explains relationships between the multiple d' measures and decision criteria. As predicted by SDT, the discriminability of perpetrator-presence versus perpetrator-absence and the discriminability of the perpetrator from an innocent suspect are relatively stable across eyewitnesses' confidence levels. However, different d' measures involving fillers make varying contributions to d'(GI) over different confidence levels. Eyewitnesses with high confidence produce good d'(GI) by their good discriminability of a guilty suspect from fillers, whereas those with low confidence show the same d'(GI) by their poor discriminability of an innocent suspect from fillers. Given that pairs of guilty and innocent suspect ID rates over different confidence levels are data points on the ROC curve, the varying contributions of discriminability measures involving fillers to d'(GI) would be reflected in producing ROC curves.

Lineup Presentation Modes in the Multi-d' Model

The multi-d' model is useful for comparing the underlying processes of eyewitness performance in simultaneous and sequential lineups. Our analyses demonstrated that d'(GI) was higher in sequential lineups for innocent-suspect studies; but higher in simultaneous lineups for average-filler studies. In innocent-suspect studies which used relatively biased lineups, sequential lineups produced a slightly lower $d'(GF_p)$ and substantially lower $d'(IF_a)$, which yielded higher d'(GI) in sequential lineups. However, the superiority of sequential lineups in innocent-suspect studies disappeared in average-filler studies which used fairer lineups. For average-filler studies, the advantage of $d'(GF_p)$ in simultaneous lineups (vs. sequential lineups) increased; and the average-filler practice led to lower $d'(IF_a)$ in simultaneous lineups, which consequently yielded higher d'(GI) in simultaneous lineups. The greater d'(GI) in sequential lineups for innocent-suspect studies is consistent with prior studies showing the superiority of sequential lineups when using biased lineups (Carlson et al., 2008; Lindsay et al., 1991). Therefore, these findings indicate that the proper use of lineup modes would depend on lineup bias—determining the bias transition point between simultaneous and sequential lineups will require further research. Preferences for one procedure or the other will depend on the bias levels reflected in actual lineups and the weights policymakers give to the various outcomes from these procedures.

Limitations and Future Directions

We note that the multi-d' model is based on simplifying assumptions which may not yield precise estimates or specifications of values. In the end our enterprise is more a series of thought experiments in which we try to identify the impact of various study characteristics (some manipulated and some not) on measures which reflect different aspects of witness performance and we ultimately rely on the relative sizes (and direction of change) in our measures to suggest conclusions. We will leave it to modelers to supply precision where that is desirable.

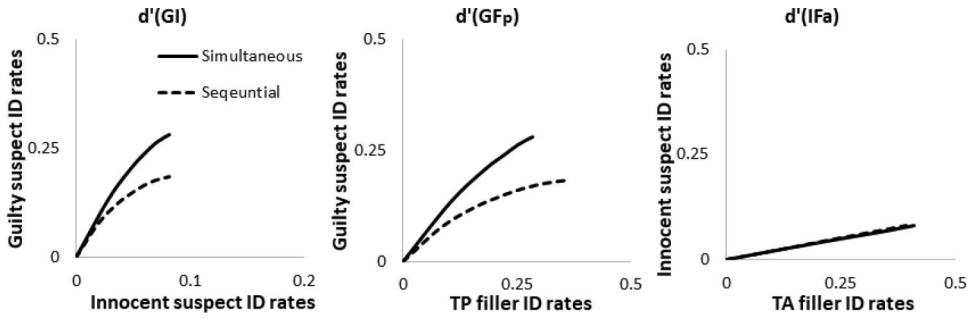


Figure 12. The application of the multi-d' model to ROC curves (data simulation: $M_{guilty} = 1$, $SD_{guilty} = 1.22$, $M_{innocent\ \&\ filler} = 0$, $SD_{innocent\ \&\ filler} = 1$, $c_{simultaneous}$ & $c_{sequential}$ calculated from Study 2 of Mickes et al. (2012), total 60,000 simulations for each lineup type).

In the present study, we applied the multi-d' model to current issues in eyewitness research. However, there remain other critical issues that may be addressed with the model. For example, according to the model, eyewitness performance in lineups may be measured by multiple ROC curves. Although ROC curves in prior eyewitness studies only focused on pairs of guilty and innocent suspect ID rates (i.e., $d'(\mathrm{GI})$), the multi-d' model suggests that eyewitness ROC curves could be also drawn with pairs of guilty suspect and TP filler ID rates (i.e., $d'(\mathrm{GF_p})$); pairs of innocent suspect and TA filler ID rates (i.e., $d'(\mathrm{IF_a})$); pairs of TP and TA filler ID rates (i.e., $d'(\mathrm{F_aF_p})$); and pairs of TP and TA rejection rates (i.e., $d'(\mathrm{R_aR_p})$). For example, plotting ROC curves for each of the d' measures can give a clearer picture of "where the action is" when parameters values change. In Figure 12, we see that a simultaneous advantage reported by Mickes, Flowe, and Wixted (2012) is associated with superior $d'(\mathrm{GF_p})$ and not with $d'(\mathrm{IF_a})$. It would be worth investigating relationships among areas under the multiple ROC curves and developing a multidimensional ROC curve (or a ROC volume), which is a single index incorporating all the multiple d' measures.

We suggest that the multi-d' model may be applied to other types of compound decision tasks as well as eyewitness identifications in lineups. There are tasks that require both detection and identification decisions simultaneously in fields of engineering, medicine, education, and so on. For example, imagine that there are two different medical tests to make a diagnostic decision of whether a patient has Disease X when there are several diseases which produce symptoms similar to Disease X. Given that this task involves a detection decision (i.e., detecting the presence of a disease) plus identification decision (i.e., discriminating Disease X from other diseases having similar symptoms), the multi-d' model may be applicable to compare diagnostic performance between the two medical tests. We hope that the multi-d' model will be a useful tool to understand decisionmakers' performance for a variety of compound decision tasks.

References

Alexander, J. R. M. (2006). *An approximation to d′ for n-alternative forced choice*. Retrieved from https://eprints.utas.edu.au/475/1/nAFCrev207.pdf

Bergold, A. N., & Heaton, P. (2018). Does filler database size influence identification accuracy? *Law and Human Behavior, 42,* 227–243. http://dx.doi.org/10.1037/lhb0000289

Bruer, K. C., Fitzgerald, R. J., Therrien, N. M., & Price, H. L. (2015). Line-up member similarity influences the effectiveness of a salient rejection option for eyewitnesses. *Psychiatry, Psychology and Law, 22,* 124–133. http://dx.doi.org/10.1080/13218719.2014.919688

Carlson, C. A., & Carlson, M. A. (2014). An evaluation of perpetrator distinctiveness, weapon presence, and lineup presentation using ROC analysis. *Journal of Applied Research in Memory & Cognition, 3,* 45–53. http://dx.doi.org/10.1016/j.jarmac.2014.03.004

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. http://dx.doi.org/10.1037/1076-898X.14.2.118

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 32,* 187–218. http://dx.doi.org/10.1007/s10979-006-9082-4

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27,* 1227–1239. http://dx.doi.org/10.1177/0956797616655789

Cutler, B. L., & Penrod, S. D. (1989). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281–290. http://dx.doi.org/10.1037/0021-9010.73.2.281

Doob, A. N., & Kirshenbaum, H. M. (1973). The effects on arousal of frustration and aggressive films. *Journal of Experimental Social Psychology, 9,* 57–64. http://dx.doi.org/10.1016/0022-1031(73)90062-0

Duncan, M. (2006). *A signal detection model of compound decision tasks* (Technical Note DRDC TR 2006–256). Ottawa, Ontario, Canada: Defence Research and Development Canada.

Evelo, A., Lee, J., Modjadidi, K., & Penrod, S. D. (2018, June). *The role of lineup bias in witness accuracy, the confidence-accuracy relationship and the courtroom value of witness confidence*. Paper presented at the Annual Conference of the European Association of Psychology and Law, Turku, Finland.

Fitzgerald, R. J., Oriet, C., & Price, H. L. (2015). Suspect filler similarity in eyewitness lineups: A literature review and a novel methodology. *Law and Human Behavior, 39,* 62–74. http://dx.doi.org/10.1037/lhb0000095

Fitzgerald, R. J., Whiting, B. F., Therrien, N. M., & Price, H. L. (2014). Lineup member similarity effects on children's eyewitness identification. *Applied Cognitive Psychology, 28,* 409–418. http://dx.doi.org/10.1002/acp.3012

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, UK: Wiley.

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. http://dx.doi.org/10.1037/a0015082

Horry, R., & Brewer, N. (2016). How target-lure similarity shapes confidence judgments in multiple-alternative decision tasks. *Journal of Experimental Psychology: General, 145,* 1615–1634. http://dx.doi.org/10.1037/xge0000227

Irwin, R. J., & McCarthy, D. (2013). Psychophysics: Methods and analyses of signal detection. In A. L. Kennon & P. Michael (Eds.), *Handbook of research methods in human operant behavior* (pp. 291–321). New York, NY: Plenum Press.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1304–1316. http://dx.doi.org/10.1037/0278-7393.22.5.1304

Key, K. N., Cash, D. K., Neuschatz, J. S., Price, J., Wetmore, S. A., & Gronlund, S. D. (2015). Age differences (or lack thereof) in discriminability for lineups and showups. *Psychology, Crime & Law, 21,* 871–889. http://dx.doi.org/10.1080/1068316X.2015.1054387

Key, K. N., Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Cash, D. K., & Lane, S. (2017). Line-up fairness affects postdictor validity and 'don't know' responses. *Applied Cognitive Psychology, 31,* 59–68. http://dx.doi.org/10.1002/acp.3302

Lampinen, J. M., Odegard, T. N., Blackshear, E., & Toglia, M. P. (2005). Phantom ROC. In D. T. Rosen (Ed.), *Trends in experimental psychology research* (pp. 235–267). New York, NY: Nova Science.

Lee, J., Mansour, J., & Penrod, S. D. (2019). [Validity of mock-witness paradigm for measuring lineup fairness]. Unpublished raw data.

Lee, J., & Penrod, S. D. (2019). *Facial identification: A meta-analysis of 50 years of research*. Manuscript in preparation.

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76,* 796–802. http://dx.doi.org/10.1037/0021-9010.76.6.796

Luus, C. A. E., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15,* 43–57. http://dx.doi.org/10.1007/BF01044829

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior, 5,* 299–309. http://dx.doi.org/10.1007/BF01044945

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 & Cognition, 33,* 783–792. http://dx.doi.org/10.3758/BF03193074

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. http://dx.doi.org/10.1037/a0030609

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 & Cognition, 3,* 58–62. http://dx.doi.org/10.1016/j.jarmac.2014.04.007

Mickes, L., Seale-Carlisle, T. M., Wetmore, S. A., Gronlund, S. D., Clark, S. E., Carlson, C. A., . . . Wixted, J. T. (2017). ROCs in eyewitness identification: Instructions versus confidence ratings. *Applied Cognitive Psychology, 31,* 467–477. http://dx.doi.org/10.1002/acp.3344

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review, 14,* 858–865. http://dx.doi.org/10.3758/BF03194112

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. http://dx.doi.org/10.1037/h0093923

Penrod, S. D. (2003). Eyewitness identification evidence: How well are witnesses and police performing? *Criminal Justice Magazine, Spring,* 36–47.

Pollack, I., & Norman, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science, 1,* 125–126. http://dx.doi.org/10.3758/BF03342823

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Myerson, T. (2018). Eyewitness identification performance on showups improves with an additional-opportunities instruction: Evidence for present-absent criteria discrepancy. *Law and Human Behavior, 42,* 215–226. http://dx.doi.org/10.1037/lhb0000284

Smith, A. M., Wells, G. L., Lindsay, R. C. L., & Penrod, S. D. (2017). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior, 41,* 127–145. http://dx.doi.org/10.1037/lhb0000219

Smith, A. M., Wells, G. L., Smalarz, L., & Lampinen, J. M. (2018). Increasing the similarity of lineup fillers to the suspect improves the applied value of lineups without improving memory performance: Commentary on Colloff, Wade, and Strange (2016). *Psychological Science, 29,* 1548–1551. http://dx.doi.org/10.1177/0956797617698528

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language, 70,* 36–52. http://dx.doi.org/10.1016/j.jml.2013.09.005

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27,* 523–540. http://dx.doi.org/10.1023/A:1025438223608

Steblay, N. K., Dysart, J. E., & 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. http://dx.doi.org/10.1037/a0021650

Swets, J. A. (1973). The relative operating characteristic in psychology: A technique for isolating effects of response bias finds wide use in the study of perception and cognition. *Science, 182,* 990–1000. http://dx.doi.org/10.1126/science.182.4116.990

Swets, J. A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems: Methods from signal detection theory*. New York, NY: Academic Press.

Tredoux, C. G. (1998). Statistical inference on measures of lineup fairness. *Law and Human Behavior, 22,* 217–237. http://dx.doi.org/10.1023/A:1025746220886

Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions? *Law and Human Behavior, 24,* 231–258. http://dx.doi.org/10.1023/A:1005463020252

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89–103. http://dx.doi.org/10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48,* 553–571. http://dx.doi.org/10.1037/0003-066X.48.5.553

Wells, G. L. (2001). Police lineups: Data, theory, and policy. *Psychology, Public Policy, and Law, 7,* 791–801. http://dx.doi.org/10.1037/1076-8971.7.4.791

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78,* 835–844. http://dx.doi.org/10.1037/0021-9010.78.5.835

Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory & Cognition, 4,* 313–317. http://dx.doi.org/10.1016/j.jarmac.2015.08.008

Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory & Cognition, 4,* 324–328. http://dx.doi.org/10.1016/j.jarmac.2015.08.010

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review, 121,* 262–276. http://dx.doi.org/10.1037/a0035940

Wixted, J. T., & Mickes, L. (2015a). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory & Cognition, 4,* 318–323. http://dx.doi.org/10.1016/j.jarmac.2015.08.009

Wixted, J. T., & Mickes, L. (2015b). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory & Cognition, 4,* 329–334. http://dx.doi.org/10.1016/j.jarmac.2015.08.007

Wixted, J. T., & Mickes, L. (2018). Theoretical vs. empirical discriminability: The application of ROC methods to eyewitness identification. *Cognitive Research: Principles and Implications*. Advance online publication. http://dx.doi.org/10.1186/s41235-018-0093-8

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018). Models of lineup memory. *Cognitive Psychology, 105,* 81–114. http://dx.doi.org/10.1016/j.cogpsych.2018.06.001

Received October 21, 2018
Revision received June 20, 2019
Accepted June 24, 2019 ■