Fair Lineups Improve Outside Observers' Discriminability, Not Eyewitnesses' Discriminability: Evidence for Differential Filler-Siphoning Using Empirical Data and the WITNESS Model

Andrew Smith
Iowa State University

Laura Smalarz
Arizona State University

Gary L. Wells
Iowa State University

James Michael Lampinen
University of Arkansas

Fair Lineups Improve Outside Observers' Discriminability, Not Eyewitnesses' Discriminability: Evidence for Differential Filler-Siphoning Using Empirical Data and the WITNESS Model

Andrew M. Smith, Laura Smalarz, Gary L. Wells, James Michael Lampinen, & Simona Mackovichova

Author Note

Andrew M. Smith, Department of Psychology, Iowa State University; Laura Smalarz, Department of Psychology, Arizona State University; Gary L. Wells, Department of Psychology, Iowa State University; James Michael Lampinen, Department of Psychology, University of Arkansas; Simona Mackovichova, Department of Psychology, Carleton University.

Correspondence concerning this article should be addressed to Andrew M. Smith, Department of Psychology, 1347 Lagomarcino Hall, Iowa State University, 50011, amsmith@iastate.edu.

This work was supported by a grant awarded to AMS from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-05336) and by a grant awarded to AMS, GLW, and JMP from Arnold Ventures (400-61-12C). The views and opinions expressed in this work are those of the authors and do not necessarily reflect the views of the granting agencies.

The authors would like to thank Melissa F. Colloff for graciously sharing her lineup materials and Andrew Evelo for sharing WITNESS code.

Abstract

Fair lineups (good fillers) better sort between innocent and guilty suspect identifications than do biased lineups (poor fillers). Why are fair lineups better? Some argue that the fair-lineup advantage is an improvement in eyewitness discriminability through some mechanism such as diagnostic-feature detection. Others argue that the fair lineups do not improve eyewitnesses' discriminability at all but instead improve the discriminability of outside observers who are privy to which lineups members are known-fillers (the differential filler-siphoning mechanism). Experiment 1 used a forced-choice paradigm to show that fair lineups do not improve eyewitness discriminability. The second experiment used the WITNESS model to show that differential filler-siphoning and the fair-lineup advantage readily surfaces and nicely patterns experimental data based on minimal assumptions even though witness memory strength was held constant. Together, these two experiments support differential filler-siphoning and the idea that fair lineups enhance the outside observer's discriminability, not the eyewitness's discriminability.

Keywords: Eyewitness Memory; Eyewitness Identification; Lineups; Signal Detection Theory; Computational Modeling

Public Significance Statement

There is consensus among psychological scientists that fair lineups lead to a better trade-off between guilty-suspect identifications and innocent-suspect identifications than do biased lineups. Yet, debate over the mechanism that leads to this fair-lineup advantage abounds. On one hand, some propose that fair lineups increase the capacity of eyewitnesses to discriminate between the guilty and the innocent. On the other hand, others have proposed that fair lineups do not increase eyewitness discriminability, but simply spread choice preferences to fillers, and that this process is more pronounced when the suspect is innocent than when the suspect is guilty. In our first experiment, we show that fair lineups do not increase perpetrator identifications even when a witness is forced to make a choice from a perpetrator-present lineup. In fact, fair lineups led to fewer perpetrator identifications, a result that is inconsistent with the improved eyewitness discriminability account for the fair-lineup advantage. In our second experiment, we used computational modeling to demonstrate that a spreading effect we call *differential filler-siphoning* produces a fair-lineup advantage even when the psychology of simulated witnesses is held constant across fair and biased lineups.

Fair Lineups Improve Outside Observers' Discriminability, Not Eyewitnesses' Discriminability: Evidence for Differential Filler-Siphoning Using Empirical Data and the WITNESS Model

Police lineups have been used to investigate criminal offenses for at least the past century (e.g., Borchard, 1932). The modern lineup is a procedure in which a person the police suspect to be the perpetrator is surrounded with known-innocent fillers (lineup lures) and shown to an eyewitness for an identification attempt. There are two basic assumptions underlying the logic of a police lineup. The first is that the eyewitness to the crime has some memory of the perpetrator that will allow her to recognize the perpetrator if she sees her again. If the person the police suspect of committing the crime is in fact the perpetrator, then the eyewitness should be able to recognize and identify that individual. Of course, recognition memory is imperfect, so the eyewitness will not always be able to recognize a perpetrator who is in the lineup, but over the long haul, the perpetrator should tend to stand out from the known-innocent fillers. The second basic assumption underlying the logic of a police lineup is that if the person who the police suspect of committing the crime is *not* the perpetrator, then the eyewitness should have no reason to recognize (let alone identify) that person. Indeed, the essential function of lineup fillers is to prevent eyewitnesses from knowing which person the police suspect, thereby precluding identifications made based on anything other than eyewitnesses' own recollections.

The logic of the police lineup breaks down if the lineup fillers do not prevent the eyewitness from discerning which lineup member is the suspect. A typical biased lineup is one in which the suspect fits the verbal description that the eyewitness gave of the perpetrator whereas the fillers do not (Wells & Luus, 1990; Wells et al., 1998; Wells et al., 2020). Suppose, for example, the eyewitness described the perpetrator as a white male, mid 20s, clean shaven, short dark hair, with a thin face. If the suspect fits that description but the fillers do not, then the suspect stands out as the obvious choice even if the suspect is innocent because the suspect is likely to better match the eyewitness's memory than any of the fillers. Research confirms that fair lineups, which use fillers that prevent the suspect from standing out (e.g., every lineup member fits the description), result in better discrimination between perpetrator identifications and innocent-suspect identifications than do biased lineups, which use fillers that let the suspect stand out (e.g., Clark, 2012; Fitzgerald et al., 2013; Lindsay & Wells, 1980).

Despite scientific consensus that law enforcement should use fair rather than biased lineups, there is debate about *why* fair lineups are superior to biased lineups. Some researchers have attributed the benefits of fair lineups over biased lineups to an improvement in eyewitness discriminability (e.g., Wetmore et al., 2015; Wixted & Mickes, 2014, 2015b). For example, the diagnostic-feature detection hypothesis proposes that fair lineups lead witnesses to appreciate that facial features that are shared across lineup members are non-diagnostic, to discount those non-diagnostic features, and to place more weight on more diagnostic facial features (Colloff et al., 2016; Wetmore et al., 2015; Wixted & Mickes, 2014). The result of decreasing the weight attached to non-diagnostic features and increasing the weight attached to diagnostic features *would* be an increase in eyewitness discriminability. In this type of account, the fair lineup is presumed to *improve the ability of the eyewitness* to better detect differences between the innocent and the guilty.

Others, including the current authors, have made the case that fair lineups do not improve the ability of the eyewitness to sort innocent from guilty at all but instead simply spread eyewitnesses' identification errors to the less forensically harmful mistake of landing on known-innocent fillers (e.g., Lee & Penrod, 2019; Smith et al., 2017; Smith et al., 2018; Smith et al., 2020; Smith et al., 2020; Wells et al., 2015; Wells et al., 2015). In this account, it is not eyewitness discriminability that improves with a fair lineup. Instead, it is the ability of an outside observer (the investigator) to sort between innocent and guilty suspect identifications that characterizes the fair lineup advantage. This improvement in the discriminability of the outside observer stems from the fact that the observer is privy to which lineup members are fillers and which one is the suspect. Smith et al. (2020) called this "investigator discriminability."

Why does an outside observer's discriminability improve with a fair lineup? It improves because good fillers (the hallmark of a fair lineup) will lure eyewitnesses' choice preferences away from an innocent suspect more than they will lure eyewitnesses' choice preferences away from a guilty suspect, a process known as *differential filler siphoning* (e.g., Lee & Penrod, 2019; Smith et al., 2017; Smith et al., 2018; Smith et al., 2020; Wells et al., 2015; Wells et al., 2015). In the differential filler-siphoning account, eyewitnesses make as many or more mistakes (misses and false positives) with fair lineups as they do with biased lineups but the false positives with fair lineups do not load up on the suspect like they do with a biased lineup and instead spread out across the fillers.

Consider the data in Table 1, which show the typical pattern in experiments that compare fair with biased lineups. These data are from an experiment in which the individual who perpetrated a simulated crime possessed a distinctive feature (e.g., a large face tattoo, black eye, etc.) (Colloff et al., 2016). After viewing the simulated-crime video, participants were presented with either a 6-person perpetrator-present lineup or a 6-person perpetrator-absent lineup. The lineup was either biased (only the suspect possessed the distinctive feature) or fair (the lineup photos were digitally altered so that the suspect did not stand out due to distinctive features).[1]

[1] The researchers used three methods for creating fair lineups: (1) photoshopping the distinctive feature so that it appeared on all lineup members and not only on the suspect; (2) covering the distinctive feature with a black block and covering the same portion of all filler faces with an identical black block; or (3) pixelating the distinctive feature and pixelating the same portion of all filler faces. The authors found no differences across these three fair lineups.

Based solely on the consideration of the relative suspect identification rates (31% hits and 9% false alarms on innocent suspect for fair; 57% hits and 36% false alarms on innocent suspect for biased), the authors concluded that fair lineups led to better eyewitness discriminability (d' = 0.84) than did biased lineups (d' = 0.53) (Colloff et al., 2016).

We agree that the d' values reported by the authors suggest that the fair lineup produced a better trade-off between guilty-suspect identifications and innocent-suspect identifications than did the biased lineup. What we contest is the inference that fair lineups increased eyewitness discriminability. In fact, an inspection of the pattern of eyewitness' identifications suggests the opposite (Table 1). The fair lineup resulted in fewer perpetrator identifications than did the biased lineup (31% vs. 57%) and about the same number of correct rejections (45% vs. 42%). In other words, the witnesses who encountered fair lineups made *fewer* correct decisions when the perpetrator was present and about the same number of correct decisions when the perpetrator was absent compared to witnesses who encountered biased lineups. That pattern suggests that fair lineups led to *worse* (not better) eyewitness discriminability than did biased lineups. The data pattern in Table 1 emerges consistently when fair lineups are compared to biased lineups (e.g., Clark, 2012; Fitzgerald et al., 2013).

Accordingly, we discuss only the replication version in the main body of the present paper and simply refer to it as a fair lineup.

Table 1

Comparison of Fair and Biased Lineups from Colloff et al. (2016)

	Perpetrator Present			Perpetrator Absent		
	Hit on Perpetrator	False Alarm on Filler	False Rejection	False Alarm on Innocent Suspect	False Alarm on Filler	Correct Rejection
Fair Lineup	31%	34%	35%	9%	46%	45%
Biased Lineup	57%	19%	24%	36%	22%	42%

This creates something of a paradox. How is it that fair lineups result in worse eyewitness discriminability than do biased lineups, and yet, produce a better trade-off between guilty-suspect identifications and innocent-suspect identifications? The key to resolving this paradox is to appreciate that there is a distinction between eyewitness discriminability and investigator discriminability (Smith et al., 2020; Wells & Luus, 1990). Unlike the eyewitness, the investigator knows which lineup member is the suspect and which lineup members are fillers. A filler identification is a *known error* and although the eyewitness did not know that fact the investigator knows that.

These data are typical in showing that fair perpetrator-absent lineups produce about the same rate of mistaken identifications (55% in Table 1) as do biased lineups (58% in Table 1). Hence, correct rejections in perpetrator-absent lineups were similar across fair and biased lineups. But the investigator is an outside observer who is privy to which lineup members are mere fillers. As a result, the investigator can ignore all the mistaken identifications of fillers. There were stark differences in how mistaken identifications were *distributed* in fair versus biased perpetrator-absent lineups. In the fair lineup five (the number of fillers) of every six (the number of lineup members) mistaken identifications are known by the outside observer to be mere filler identifications. In effect, the fillers in the fair lineup siphoned off most of the mistaken identifications into the known-error category (46% of the 55%) and only a fraction of the mistaken identifications landed on the innocent suspect (9% of the 55%). Conversely, the fillers in the biased lineup siphoned only about 1/3rd of the mistaken identifications (22% of the 58%) and nearly 2/3rds of all mistaken identifications landed on the innocent suspect (36% of 58%). Hence, an outside observer (e.g., the investigator) who knows which lineup members are fillers is only at risk of missing 9% of mistaken identifications in the fair lineup (which are unknown errors) whereas an outside observer is at risk of missing 36% of mistaken identifications in the biased lineup (four times as many unknown errors). The eyewitnesses are not better at avoiding false identifications with a fair perpetrator-absent lineup; it is simply the case that more of their mistaken identifications are distributed to fillers when the lineup is fair rather than biased. The fair lineup improves the outside observer's ability to sort between innocent and guilty suspects, but it does not do this by improving the ability of the eyewitness at all.

We have broken down fair versus biased lineup this way before (e.g., Smith et al., 2018; Wells et al., 2015) because it helps to clarify that fair fillers do not improve discriminability for the eyewitness but instead only improves discriminability for the investigator. Notice that fair fillers also siphon some choices away from the guilty suspect in a perpetrator-present lineup. However, fair fillers siphon more from the innocent suspect (in perpetrator-absent) than from the guilty suspect (in perpetrator-present), which is what is meant by *differential* filler-siphoning. It is because filler siphoning is *differential* that fair lineups improve the capacity of the investigator to discriminate guilty-suspect identifications from innocent-suspect identifications (Smith et al., 2017). It makes sense that fair-lineup fillers would siphon more from the innocent suspect than from the guilty suspect because in a fair lineup the fillers and the innocent suspect are drawn from the same similarity distribution and there are typically five such fillers (vs. only one innocent suspect). Hence, these fillers in a fair perpetrator-absent lineup are effective in drawing the eyewitnesses' preferences away from the innocent suspect. These same five fillers in a perpetrator-present lineup, in contrast, should draw fewer preferences away from the perpetrator because the witness has seen the perpetrator before (at the scene of the crime).

The result of this differential filler-siphoning process is that fair fillers decrease innocent-suspect identifications to a greater extent than guilty-suspect identifications, thereby increasing the capacity of investigators to discriminate guilty-suspect identifications from innocent-suspect identifications ($d'_{\text{Fair}} = 0.84$ vs $d'_{\text{Biased}} = 0.53$). Importantly, however, notice how this d' is characterizing the ability of the *outside observer* (investigator), who is privy to which lineup members are fillers, to better sort between innocent and guilty suspects in fair versus biased lineups; *the ability of the eyewitness to identify the perpetrator and avoid false positives has not been improved at all by the fair lineup compared to the biased lineup.*

Research Overview of Experiments 1 and 2.

Despite what we think are strong arguments and evidence since at least 2015 supporting the differential filler-siphoning interpretation of the fair lineup advantage, there continues to be speculation that some other mechanism in fair lineups is operating (such as the diagnostic feature detection hypothesis) that improves eyewitness discriminability. So, in our first experiment, which uses a six-alternative forced-choice lineup test, we offer the most direct and simple test yet of the proposition that fair lineups improve eyewitness discriminability. If fair lineups increase eyewitness discriminability, then with a forced-choice perpetrator-present test fair lineups should lead to more perpetrator identifications than biased lineups. Not surprisingly, fair lineups led to fewer perpetrator identifications. There is no way to reconcile that pattern of results with the idea that fair lineups increase eyewitness discriminability. We contrast this with a manipulation of memory strength (1 versus 3 opportunities to watch the encoding video); not surprisingly, the strong memory condition did increase the perpetrator identification rate. In our second experiment, we use the WITNESS model (Clark, 2003) to show that differential filler siphoning, and the fair-lineup advantage are easily produced and closely replicate actual data without any need to assume an increase in eyewitness discriminability. This evidence from the WITNESS model is important because several researchers have claimed that they could not reproduce differential filler-siphoning in their computational models. We explain where their models went wrong and show that the problem was much simpler to resolve than even we expected it to be.

Experiment 1: Using a Modified Forced-Choice Lineup to Test the Prediction that Fair Lineups Improve Eyewitness Discriminability

Our first experiment used materials from Colloff et al. (2016) to replicate the fair-lineup advantage. Importantly, we also included a second step in the procedure; we had witnesses who initially rejected the lineup treat the lineup as a six-alternative forced-choice task. The meaning of the forced-choice data in the perpetrator-absent conditions is unclear and not of any particular interest. But the meaning of the forced-choice data in the perpetrator-present lineup constitutes a simple, direct test of whether a fair lineup helps witnesses discriminate between guilty and innocent. The forced-choice perpetrator-present portion is a simple test because there is no need to statistically control for decision criterion (everyone chooses). The forced-choice perpetrator-present portion is a clean, direct test because it is a simple array of the guilty person embedded among innocents. If fair lineups help eyewitnesses discriminate between the guilty and the innocent (e.g., as the diagnostic feature detection hypothesis posits), then forced-choice accuracy in the perpetrator-present condition should be higher for the fair lineup than for the biased lineup.

Methods

The Research Ethics Board at Carleton University approved this experiment.

Participants

Participants were not permitted to use a cell phone to complete this study; those participants who attempted to do so were identified during the consent process, booted from the study at that time, and asked to complete the study using either a tablet or computer. In total, 1734 Qualtrics workers participated in this experiment in exchange for points that could be used to buy gift certificates. The value of this remuneration was approximately $1.00 USD. We excluded 368 participants per our pre-registered experiment plan: 226 participants failed a simple "bot" check (e.g., "please select 'strongly agree' from the list below), 48 participants who experienced a technical issue with the encoding video, five participants who experienced a technical issue with the lineup photographs, 23 participants who indicated that the stimuli appeared familiar from a previous study, and 66 participants who failed to indicate what happened during the encoding video. Many participants failed multiple exclusion criteria, but we excluded participants for each criterion in the order in which we present them.

Our final sample included 1365 Qualtrics workers who were an average of 54.84 years old (*SD* = 15.79). Half of the participants were female (female = 50.48%; male = 49.44%, and other = 0.07%). Eighty-two percent of participants identified as White or Caucasian, 9% identified as Black or African American, 4% identified as Hispanic or Latino/Latina, 3% identified as Asian, 2% identified as other, and 1% identified as American Indian or Alaska Native.

Design

Each participant was randomly assigned to one cell of a 2 (strong memory, weak memory) × 2 (fair lineup, biased lineup) × 2 (perpetrator, innocent-suspect) between-participants design.

Materials

Perpetrator Videos. We used three of the four encoding videos that were initially used in Colloff et al. (2016): the carjacking scenario, the graffiti scenario, and the mugging scenario (we omitted the theft scenario because the distinctive feature – a nose piercing – was small and we feared that many participants would likely fail to notice it). In the carjacking scenario, a white male in his mid-20s with a large scar on his cheek opened the driver-side door, yelled at the female to exit the vehicle, climbed into the car, went through the victim's belongings, and then drove off. In the graffiti scenario, a white male in his early 20s with a large black eye walked into a classroom and spray painted "Uni sucks" on the wall. In the mugging scenario, a white male in his early 20s with a facial tattoo on his right cheek approached a male victim, stole his cell phone, and ran off with it while the victim gave chase. Each video was approximately 30 seconds in length.

Anagrams Task. After watching the encoding video, participants completed a 6-minute word unscrambling (anagrams) task. The purpose of this task was to increase the memory-retention interval.

Lineup Manipulations. After completing the anagrams task, participants were randomly assigned to view either a perpetrator-present or perpetrator-absent fair or biased lineup. We used the same lineup stimuli as Colloff et al. (2016). We programmed Qualtrics survey platform to randomize the lineup stimuli for each individual participant. In the perpetrator-present conditions, the perpetrator was surrounded by five fillers who were randomly drawn from a larger pool of 40 fillers (Colloff et al., 2016). In perpetrator-absent conditions, six fillers were randomly drawn from the larger pool of 40 fillers.

In the fair conditions, each of the lineup fillers possessed the same distinctive feature as the perpetrator (using photoshop, Colloff et al. (2016) replicated the perpetrator's distinctive feature and placed it on each of the lineup fillers) and the innocent-suspect identification rate for the perpetrator-absent lineup was estimated by dividing the total false-positive rate by the total number of lineup members (six). In the biased conditions, only the suspect possessed the distinctive feature; the fillers did not. For biased perpetrator-absent lineups, we used a designated innocent suspect. One filler who possessed the distinctive feature was drawn from the pool of 40 photos and five fillers who did not possess the distinctive feature were drawn from the remaining 39 photos in the pool. The filler who possessed the distinctive feature served as the designated innocent suspect. Finally, across all conditions, Qualtrics survey platform randomized the positions of lineup members. Hence, every participant in this experiment viewed a unique lineup configuration.

Procedure

After providing informed consent, participants were asked to indicate their age, gender, and race/ethnicity. Participants were then instructed that they would watch a 30-second video and to pay careful attention as they would be asked questions about the video later. Participants in the strong-memory condition viewed the encoding video a total of three times, and prior to the second and third viewing were directed to pay attention to the perpetrator as we would test their memory for him later in the experiment. Participants in the weak-memory condition viewed the encoding video only once. After watching the encoding video, all participants completed the 6-

minute anagrams task. Participants were then provided with the following lineup instructions: *On the next page you will view a lineup that may or may not include the man you saw in the video at the beginning of the experimental session. If you believe that the man from the video is present in the lineup, you should identify that person. If you do not believe that the man from the video is present in the lineup, you should select "not present".*

Participants then viewed the lineup, above which the lineup instructions were again displayed. Participants who affirmatively identified someone from the lineup were asked to indicate how confident they were in their decision from 0% to 100% in 10-point increments. Participants who rejected the lineup were not asked for their confidence and instead were given the following instructions: *You selected the "not there" option, but what we would like you to do now is to identify the person who best matches your memory for the man you saw in the video.* After selecting the lineup member who best matched their memory for the perpetrator, these participants were asked to indicate how confident they were that this lineup member was the perpetrator.

After completing the lineup procedure and providing a confidence statement, participants answered a series of quality-control questions. Participants were asked to indicate what happened in the video and were provided with six options, one of which was a correct description of what they witnessed. Participants were then asked to select "Strongly Agree" from a list of options on a 5-point Likert scale and were asked if they had seen the video and/or photographs prior to taking part in this study (e.g., in a different study). Participants were then asked to indicate if they encountered any technical issues with the video or the lineup photos. After answering these quality-control questions, participants were fully debriefed, thanked, and compensated for their time.

Results

Our preregistration and analysis plan included several hypotheses and supplemental analyses (e.g., that increasing memory strength would increase perpetrator identifications and decrease false alarms). For brevity, in the main body of this paper, we focus only on those analyses that most directly test the increased eyewitness discriminability account for the fair-lineup advantage. Remaining analyses and hypothesis tests are available in online supplemental materials. We used RStudio (RStudio Team, 2015), Tidyverse (Wickham, 2017), and gridExtra (Baptiste, 2017) to analyze our data.

Table 2 displays the proportions of eyewitnesses who identified the perpetrator, a filler, an innocent suspect, or made no identification as functions of memory strength, suspect guilt, and lineup fairness. After making an initial identification decision, participants who indicated that the perpetrator was not present were then asked to indicate which lineup member best matched their memory for the perpetrator. In essence, this turned the lineup procedure into a forced-choice task. Table 3 provides the final forced-choice suspect identification and filler identification rates that includes both initial suspect and filler identifications plus suspect and filler identifications that occurred after an initial not-present response.

Table 2

Initial Eyewitness Identification Decisions as Functions of Suspect Guilt, Memory Strength, and Lineup Fairness

		Perpetrator Present			Perpetrator Absent		
		Hit on Perpetrator	False Alarm on Filler	False Rejection	False Alarm on Innocent Suspect	False Alarm on Filler	Correct Rejection
Strong Memory	Fair Lineup	37% (63)	25% (43)	38% (64)	8% (14.5)	40% (72.5)	52% (93)
	Biased Lineup	72% (133)	5% (10)	22% (41)	40% (76)	12% (22)	49% (93)
Weak Memory	Fair Lineup	30% (49)	35% (57)	35% (56)	10% (8.4)	50% (75.6)	40% (56)
	Biased Lineup	65% (104)	8% (13)	26% (42)	47% (85)	18% (33)	34% (61)

Table 3

Forced-Choice Suspect and Filler Identification Rates as Functions of Suspect Guilt, Memory Strength, and Lineup Fairness

		Perpetrator Present		Perpetrator Absent	
		Hit on Perpetrator	False Alarm on Filler	False Alarm on Innocent Suspect	False Alarm on Filler
Strong Memory	Fair Lineup	52% (89)	48% (81)	17% (30)	83% (150)
	Biased Lineup	85% (157)	15% (27)	62% (119)	38% (72)
Weak Memory	Fair Lineup	41% (67)	59% (95)	17% (23.3)	83% (116.7)
	Biased Lineup	82% (131)	18% (28)	68% (122)	32% (57)

Note. The values in this table include both initial identifications of perpetrators, innocent suspects, and fillers, plus perpetrator, innocent suspect, and filler identifications that occurred after an initial not-present response.

We used a probit regression analysis to provide a direct test of the improved eyewitness discriminability account for the fair-lineup advantage. If the improved eyewitness discriminability account is correct, then fair lineups should lead to more perpetrator identifications than biased lineups. Because only perpetrator-present conditions are relevant to testing this hypothesis, we included only data from perpetrator-present conditions in this analysis. We treated both free-choice perpetrator identifications and forced-choice selections of the perpetrator (following an initial rejection) as perpetrator choices and coded these behaviors as 1s and all other outcomes as 0s. Inconsistent with the diagnostic-feature-detection hypothesis, however, participants who viewed a fair lineup were less likely to identify the perpetrator than were participants who viewed a biased lineup, $B = -1.07$, $SE = 0.11$, $Z = -10.00$, $p < .001$. This suggests that the fair lineup did not improve eyewitness memory performance relative to the biased lineup. But this was not because it was not possible to improve eyewitness memory performance: Participants in the strong memory condition were more likely to choose the perpetrator than were participants in the weak memory condition, $B = 0.21$, $SE = 0.11$, $Z = -2.01$, $p = .045$. Hence, increasing memory strength improved eyewitness memory performance, but fair lineup fillers impaired memory performance.

For completeness, we also used a complimentary model to examine innocent-suspect identification decisions for the forced-choice task. Not surprisingly, biased lineups led to more innocent-suspect choices than did fair lineups, $B = 1.72$, $SE = 0.12$, $Z = -14.47$, $p < .001$. This finding makes sense because in the biased lineups, the innocent-suspect stood out from the other lineup members because of his distinctive feature, whereas in the fair lineups, he did not. The number of encoding repetitions did not significantly impact innocent-suspect choices, $B = 0.14$, $SE = 0.11$, $Z = 1.27$, $p = .21$.

Contrary to the diagnostic-feature-detection hypothesis, the fair lineup did not improve eyewitness discriminability on the perpetrator-present forced-choice lineup. It is very clear what improved eyewitness discriminability should look like in a six-alternative forced-choice task. In fact, these data show that improved eyewitness discriminability when we compare the strong versus weak memory conditions. That improvement in discriminability that occurs when comparing a strong versus weak memory is not happening when comparing a fair versus biased lineup. In fact, fair lineups *reduced* eyewitness discriminability. There is no way to reconcile that pattern of results with any notion that the fair-lineup advantage is attributable to an increase in eyewitness discriminability. Fair lineups simply do not lead to better eyewitness discriminability than do biased lineups.

Experiment 2: Using the WITNESS Model to Examine a Fair-Lineup Advantage

The purpose of Experiment 2 was to test whether a simple computational model that holds witness memory strength constant can account for differential filler-siphoning and produce the well-documented pattern of the fair lineup advantage. Proponents of the diagnostic-feature detection account have long rejected differential filler-siphoning as an explanation for the fair-lineup advantage on the basis that their models did not support the account (e.g., Colloff et al., 2016; Wixted & Mickes, 2015b). However, the models that these authors used suffered from a common and fatal flaw. Namely, these models assumed statistical independence. But the assumption of statistical independence is unrealistic in either the lab or the real world. Specifically, the probability of being identified is more similar for members *within* lineups than for *members* between lineups. After all, different witnesses adopt different decision criteria and where a witness places her criterion affects the probability of identification for all members in that lineup (Smith et al., 2017). Subsequently, Wetmore et al. (2017) demonstrated that variation in encoding strength also contributes to these statistical dependencies. Both criterial variance and encoding variance are methods for modeling dependencies and when these dependencies are included in the model *differential* filler-siphoning and the fair-lineup advantage emerge.

More recently others have rejected the assumption of independence on the basis that lineup members are matched to the perpetrator on a common set of descriptors, and therefore the signal strengths emanating from members of the same lineup should be more similar to each other than they are to members of other lineups (Akan et al., 2020; Wixted et al., 2018). For example, if an investigator were to implement a match-to-description strategy for constructing a lineup, all innocent lineup members would be selected, in part, because they matched the witness' description of the perpetrator. The implication is that, when a witness provides a particularly detailed description of the perpetrator, all innocent lineup members would match the perpetrator on several features, and when a witness provides a less detailed description, all innocent lineup members would match the perpetrator on few features. Hence, we should expect less variability in the signal strengths emanating from innocent members within the same lineup than we should expect between members of different lineups.

In the present experiment, we instantiated WITNESS to simulate the lineup construction process described in the preceding paragraph. We assume all innocent members in a fair lineup (innocent suspect, perpetrator-absent fillers, perpetrator-present fillers) match the perpetrator (and each other) on the same common set of features. Variation in the signal strengths of innocent persons embedded in the same lineup is solely attributable to random noise (or coincidence). This is the first time that WITNESS has been instantiated in a manner that truly reflects the lineup construction process and in a manner that naturally leads to dependent (or correlated) signal strengths. This is also the first time any computational model has compared fair and biased lineups under the assumption of dependent (cf. independent) signal strengths. Under these conditions, we show that both differential filler-siphoning and the fair lineup advantage surface without any need to resort to assumptions about differences in eyewitness discriminability. We show this to be the case for both the strong and weak memory conditions from Experiment 1 and for the fair and biased lineups from Colloff et al. (2016).[2] We offer technical discussion of why the magnitude of both differential filler-siphoning and the fair-lineup advantage increases when the independence assumption is relaxed in our supplemental materials.

Method

WITNESS is a direct-access matching-model in which recognition decisions are based on the match between stimulus probes and a memory trace (Clark & Gronlund, 1996). In the context of a lineup, this means that identification decisions are based on the match between an eyewitness' memory trace and the lineup members. The WITNESS architecture makes the following representational assumptions: (1) all faces (perpetrator, innocent suspect, fillers) are represented as vectors of 100 features, with each feature sampled from a uniform distribution with a range from -.5 to .5. (2) The face of a perpetrator is constructed by randomly sampling 100 features from the previously specified uniform distribution. (3) The WITNESS encodes the face of the perpetrator imperfectly, so that only s ($0 < s < 1$) features are correctly encoded. Correctly encoded features are assumed to perfectly match the perpetrator and are stored in a memory vector. The $1 – s$ incorrectly-encoded features are randomly sampled from the same uniform distribution as the correctly encoded features. (4) The innocent suspect and lineup fillers bear some resemblance to the perpetrator but are not identical. (5) During the lineup task, all lineup members are compared to memory by computing the dot product between each face in the lineup and the memory trace. The lineup member associated with the highest dot product provides the best match-to-memory. (6) If the memory evidence surpasses the decision criterion,

[2] In fact, we found evidence of differential filler-siphoning and a fair lineup advantage even when we assumed independent signal strengths (cf. Wetmore et al., 2017; cf. Wixted & Mickes, 2015; cf. Colloff et al., 2016); however, under the erroneous assumption of independence, the model consistently underestimated the magnitude of the fair-lineup advantage.

WITNESS makes an identification and otherwise WITNESS rejects the lineup (Clark, 2003; Clare & Lewandowsky, 2004; Lewandowsky & Farrell, 2011).

Many of the representational assumptions of the WITNESS model require some operationalization (e.g., the decision rule). We outline our operationalizations of those assumptions here. Because we modeled both fair and biased lineups, we used two parameters to govern the similarity between the innocent lineup members and the perpetrator: similarity fair (*simf*) and similarity biased (*simb*), where $0 < simb < simf < 1$. Both the innocent suspect in the fair lineup and the innocent suspect in the biased lineup were matched to the perpetrator at *simf*. Fair fillers were matched to the perpetrator on the same features as the innocent suspect and all variance in signal strength was attributable to random variability. This mimics how investigators or experimenters build fair lineups in the real world, typically by selecting fillers who fit the verbal description that the eyewitness gave of the perpetrator. For example, if the witness described hair color, style, length, gender, age, facial hair, race, distinctive features, and so on, then not only the suspect but also every filler shares those same features in a fair lineup. Even though all innocent members are matched to the perpetrator on the same features, features that do not match the perpetrator still vary in degrees of similarity to the perpetrator, which allows for variation and for one innocent lineup member to (by chance) better resemble the perpetrator than another. Biased fillers were independently matched to the perpetrator at *simb*. Although many decision strategies have been implemented in WITNESS, we used the simplest rule, namely the best-above criterion rule (Clare & Lewandowsky, 2004). If the best-matching lineup member exceeds the decision criterion, *c*, WITNESS identifies that lineup member. If none of the lineup members exceed the decision criterion, WITNESS indicates “not present”.

We fit the WITNESS model separately to the fair and biased ROC curves from our strong and weak memory conditions in Experiment 1 and to the fair and biased ROC curves from Colloff et al. (2016). For simplicity, we binned confidence levels to create four operating points for each ROC curve: >= 90, >= 70, >= 50, and >= 0. Hence, there were four decision criteria for the fair lineup, Fair90, Fair70, Fair50, and Fair0, and four decision criteria for the biased lineup, Biased90, Biased70, Biased50, and Biased0. All three instantiations of the WITNESS model therefore entailed 11 free parameters: encoding strength (*s*), the similarity between innocent suspects and fair lineup fillers to the perpetrator (*simf*), the similarity between biased lineup fillers and the perpetrator (*simb*), and eight decision criteria.

The WITNESS model completed 10,000 loops to simulate 10,000 participants making eyewitness decisions for each condition. On each loop, the fair and biased lineups used the same perpetrator and innocent suspect, the present fair and absent fair conditions used the same fillers, and the present biased and absent biased used the same fillers. The model output provides estimates of suspect and filler identification rates made in each of our four confidence bins as well as an estimate of the overall not-present rate. We used these estimates to contrast WITNESS-predicted suspect-only ROC curves with the data collected in Experiment 1 and with the data from Colloff et al. (2016). Code for all WITNESS simulations was adapted from Lewandowsky and Farrell (2011).

It is essential to appreciate that the ONLY difference between the WITNESS-simulated fair and biased lineups was the similarity of the fillers to the perpetrator (*simf* vs. *simb*). Fair and biased lineups used the same perpetrators and innocent suspects and encoding strength (*s*) was fixed across fair and biased lineups. Hence, the model assumes no psychological differences between the witnesses who encounter fair lineups and the witnesses who encounter biased lineups. If WITNESS can account for the fair-lineup advantage under these conditions, that would provide the strongest evidence to date that the fair-lineup advantage is not attributable to a change in the psychology of the witness, but instead to an improvement in the capacity of the outside observer to discriminate guilty-suspect identifications from innocent-suspect identifications. To the extent that the fair-lineup advantage can be explained by differences between filler similarities in fair and biased lineups, there would be no need to resort to assumptions about improved eyewitness discriminability between witnesses who encounter fair lineups and witnesses who encounter biased lineups.

Results

WITNESS's parameter estimates are summarized in Table 4. Table 5 shows both the observed and predicted (in parentheses) suspect and filler identification rates at each confidence bin along with the overall not-present rates. With only slight discrepancies, the present instantiation of WITNESS was clearly able to characterize the major trends in the data for both the strong and weak memory conditions. Figure 1 shows the suspect-only ROC curves for fair and biased lineups. Triangles represent the observed operating points and circles represent WITNESS-predicted operating points. Critically, WITNESS produces the same fair-lineup advantage pattern that is observed in the empirical data. That is, the fair suspect-only ROC curve bows closer to the upper left corner of the ROC space (has a higher partial Area Under the Curve) than does the biased suspect-only ROC curve. WITNESS was able to produce this fair-lineup advantage without resorting to any assumptions about psychological differences between the witnesses who encountered fair lineups and the witnesses who encountered biased lineups.

Table 4

Parameter Estimates for Best-Fitting WITNESS Models

	Strong Memory	Weak Memory	Colloff et al.
Encoding Strength (*s*)	.29	.26	.28
Similarity Fair (*simF*)	.71	.84	.80
Similarity Biased (*simB*)	.14	.19	.40
Fair90	3.29	3.38	3.49
Fair70	2.73	2.87	2.96
Fair50	2.42	2.41	2.56
Fair0	2.32	2.06	2.19
Biased90	2.46	2.54	2.97
Biased70	2.15	2.18	2.60
Biased50	1.94	1.89	2.23
Biased0	1.84	1.72	2.00

Note. Both innocent suspects were matched to the perpetrator at *simF*. Fair fillers matched the perpetrator on the same features as the innocent suspect. Biased fillers were matched to the perpetrator, independently, at *simB*. Parameters starting with Fair or Biased are decision criteria and the number associated with each criterion parameter reflects the lower bound confidence value for that criterion. The strong and weak conditions were fit separately as was the data from Colloff et al. (2016).

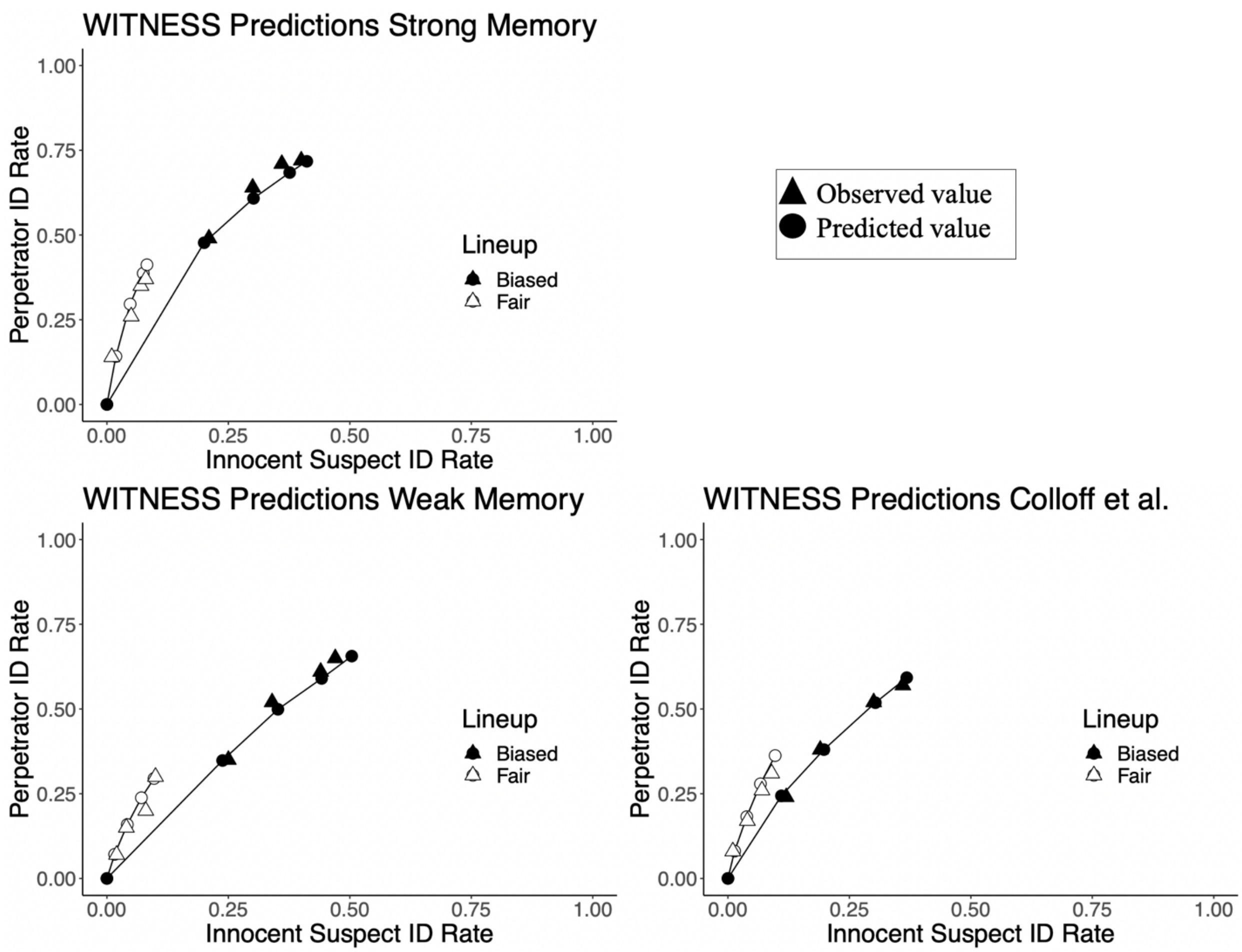
Table 5
Observed and WITNESS-Predicted Identification Rates (in parentheses) by Lineup Bias

	Strong Memory					
	Perpetrator Present			Perpetrator Absent		
	Hit perpetrator	FA Filler	False Rejection	FA Innocent Suspect	FA Filler	Correct Rejection
Fair						
90-100	.14 (.14)	.09 (.05)		.01 (.02)	.07 (.09)	
70-100	.26 (.30)	.17 (.12)		.05 (.05)	.25 (.24)	
50-100	.35 (.39)	.23 (.18)		.07 (.07)	.36 (.36)	
0-100	.37 (.41)	.25 (.20)	.38 (.39)	.08 (.08)	.40 (.40)	.52 (.52)
Biased						
90-100	.49 (.48)	.00 (.01)		.21 (.20)	.02 (.02)	
70-100	.64 (.61)	.03 (.03)		.30 (.30)	.03 (.05)	
50-100	71 (.68)	.05 (.04)		.36 (.38)	.08 (.09)	
0-100	.72 (.72)	.05 (.05)	.22 (.24)	.40 (.41)	.12 (.11)	.49 (.48)
	Weak Memory					
	Perpetrator Present			Perpetrator Absent		
	Hit perpetrator	FA Filler	False Rejection	FA Innocent Suspect	FA Filler	Correct Rejection
Fair						
90-100	.07 (.07)	.07 (.06)		.02 (.02)	.09 (.08)	
70-100	.15 (.15)	.15 (.14)		.04 (.04)	.21 (.20)	
50-100	.20 (.25)	.27 (.24)		.08 (.07)	.38 (.36)	
0-100	.30 (.32)	.35 (.33)	.35 (.35)	.10 (.10)	.50 (.50)	.40 (.40)
Biased						
90-100	.35 (.35)	.00 (.02)		.25 (.22)	.03 (.02)	
70-100	.52 (.49)	.02 (.04)		.34 (.34)	.05 (.06)	
50-100	.61 (.60)	.06 (.07)		.44 (.45)	.11 (.10)	
0-100	.65 (.66)	.08 (.09)	.26 (.25)	.47 (.51)	.18 (.14)	.34 (.35)
	Colloff et al. (2016)					
	Perpetrator Present			Perpetrator Absent		
	Hit perpetrator	FA Filler	False Rejection	FA Innocent Suspect	FA Filler	Correct Rejection
Fair						
90-100	.08 (.08)	.06 (.05)		.01 (.01)	.07 (.07)	
70-100	.17 (.18)	.14 (.12)		.04 (.04)	.19 (.20)	
50-100	.26 (.28)	.25 (.21)		.07 (.07)	.34 (.34)	
0-100	.31 (.36)	.34 (.28)	.35 (.36)	.09 (.10)	.46 (.47)	.45 (.43)
Biased						
90-100	.24 (.24)	.02 (.02)		.12 (.11)	.02 (.03)	
70-100	.38 (.38)	.06 (.05)		.19 (.20)	.07 (.09)	
50-100	.52 (.52)	.12 (.09)		.30 (.30)	.14 (.17)	
0-100	.57 (.59)	.19 (.12)	.25 (.29)	.36 (.37)	.22 (.23)	.43 (.40)

Note. FA = False Alarm. Values not in parentheses are observed values and values inside parentheses are WITNESS-predicted values.

Figure 1

Observed and WITNESS-Predicted ROC Curves as a Function of Lineup Bias



All past instantiations of the WITNESS model have tacitly assumed that which features one innocent lineup member matches the perpetrator on are independent of which features another member of the same lineup matches the perpetrator on (e.g., Clark, 2003; Clark et al., 2011). The result is that, within the same fair lineup, innocent persons might match the perpetrator on completely nonoverlapping sets of features and one innocent member might match the perpetrator on several features and another on very few features. This does not mimic the reality of what fair lineups look like in either real cases or eyewitness experiments and the assumption of independence can clearly be rejected. When investigators (or experimenters) put together a fair lineup, innocent members are chosen because they match the perpetrator on some common set of features. The result is that innocent lineup members match the perpetrator on the same features and on about the same number of features. There is undoubtedly marked variability in how similar innocent persons are to the perpetrator across fair lineups (based, for example, on differences in description quality), but within a given lineup, innocent members (both fillers and innocent suspects) should possess relatively equal degrees of similarity to the perpetrator. When WITNESS constructs lineups in this fashion that reflects how lineups are constructed in the real world, both differential filler-siphoning and the fair-lineup advantage readily surface.

General Discussion

In two experiments, one empirical and one simulated, we showed that there is no evidence to suggest that fair lineups produce better eyewitness discriminability than do biased lineups. Instead, it is the ability of the investigator (an outside observer privy to which lineup members are fillers) to discriminate between innocent and guilty suspect identifications that is enhanced by fair versus biased lineups. If fair lineups were superior to biased lineups because they increased eyewitness discriminability, then they should have resulted in more perpetrator identifications on the forced-choice lineups in Experiment 1. Likewise, if an increase in eyewitness discriminability were required to find a fair-lineup advantage (Colloff et al., 2016; Wixted & Mickes, 2015b), then we should not have found a fair-lineup advantage when implementing WITNESS in Experiment 2. Of course, we did find a fair-lineup advantage in Experiment 2. WITNESS displayed the standard fair-lineup advantage pattern found in empirical experiments even though the fair lineups led to worse eyewitness discriminability than did the biased lineups.

What is it about fair lineups that leads them to increase the capacity of an outside observer to sort between guilty- and innocent-suspect identifications relative to a biased lineup? The model-predicted values from our WITNESS simulation (displayed in Table 5) provide an answer that is very similar to the one we provided in our introduction. The model predicted fewer guilty-suspect identifications from fair than from biased perpetrator-present lineups and predicted similar rates of total false positives from fair and biased perpetrator-absent lineups. In other words, the model predictions not only failed to support the idea that fair lineups enhance eyewitness discriminability; the model-predicted values are consistent with the idea that fair lineups *decrease* eyewitness discriminability. And yet, it is clear from the simulated ROC curves in Figure 1 that fair lineups allow an outside observer to better discriminate guilty-suspect identifications from innocent-suspect identifications than do biased lineups.

The instantiation of the WITNESS model in Experiment 2 clarifies that it is not eyewitness discriminability that improves with fair lineups. Fair lineups are making eyewitness discriminability worse. In fact, as we showed in Experiment 1, it is nonsense to think that increasing the similarity of lures to a target will make people better at discriminating the target from the lures (see Smith et al., 2018). But what fair lineups do is make a large share of all the eyewitnesses' lineup preferences land on known-innocent fillers (which is not the case with the biased lineup). The outside observer, being privy to which lineup members are known-innocent fillers, can ignore the filler identifications and only needs to contend with a much lower rate of false positives on the innocent suspect.

One might reasonably question the assumption of our model that all innocent persons match the perpetrator on precisely the same features. This is both a restrictive and simplistic assumption. The innocent members of a fair lineup are all relatively likely to match the perpetrator on broad features like age, race, biological sex, hair color, etcetera, but there is likely some variance in the precise number of features and exactly what features each lineup member matches the perpetrator on. We have two responses to this line of inquiry. First, we remind the reader that there still is variance in how strongly members of the same lineup match the perpetrator. Indeed, on any feature for which innocent members are not matched to the perpetrator, there is variance in the degree of similarity between the innocent lineup members and the perpetrator. For example, if the innocent members were not matched to the perpetrator on eye color (e.g., because the eyewitness did not or could not report eye color), some would naturally provide a closer resemblance to the perpetrator on this feature than would others. Second, it is worth noting that we could add an additional parameter to our model that adds some modest degree of variability to which features and how many features each innocent person matches the perpetrator on. But even without that additional complexity, our model is providing a good account of the distinction between fair and biased lineups. Perhaps on some future dataset, the present version of the WITNESS model will not be able to account for the results without additional complexity. Until that time comes, we follow the principle of parsimony and make our model only as complex as is necessary to account for the data.

Concluding Remarks

The observation that the fair lineup advantage results from spreading false positives away from an innocent suspect and onto known-innocent fillers has a long history (Wells, 2001). In 2015, however, debate emerged over whether the differential filler-siphoning account, which is not an improvement in eyewitness discriminability, could account for why fair lineups produce a higher suspect-only ROC curve than do biased lineups or whether some other mechanism that improves eyewitness discriminability, such as diagnostic feature detection, is responsible (see

Wells et al., 2015; Wells et al., 2015; Wixted & Mickes, 2015a, 2015b). We believe that the current work largely puts that issue to rest.

Previous attempts to cleanly model differential filler siphoning all made unwarranted assumptions prolonging this debate about how lineups work. After all, if differential filler-siphoning cannot be readily modeled, then there must be some other mechanism driving an improvement in eyewitness discriminability (e.g., Colloff et al., 2016; Wixted & Mickes, 2015b). The current article shows that when WITNESS constructs lineups in a manner that mimics how lineups are constructed in the real world, both differential filler-siphoning and the fair-lineup advantage surface. Critically, this fair-lineup advantage was observed even though memory strength was fixed, and witnesses encountering fair and biased lineups were assumed behave in the same manner; namely, by identifying the best-matching lineup member if that person exceeded their decision criterion and otherwise by rejecting the lineup.

The evidence is clearer than ever that fair lineups do not result in any improvement over biased lineups in eyewitness discriminability. Instead, fair lineups produce an advantage because they distribute a large share of their lineup choice preferences away from the suspect and on to fillers, a process that is more pronounced when the suspect is innocent than when the suspect is the perpetrator. It is the outside observer, who is aware of which lineup members are merely fillers, whose discriminability is improved by using good fillers (Smith et al., 2020).

References

Akan, M., Robinson, M. M., Mickes, L., Wixted, J. T., & Benjamin, A. S. (2021). The effect of lineup size on eyewitness identification. *Journal of experimental psychology. Applied*, *27*, 369–392. doi: 10.1037/xap0000340.

Baptiste, A. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Borchard, E. B. (1932). *Convicting the Innocent: Sixty-Five Actual Errors of Criminal Justice.* New Haven, Yale University Press.

Clare, J., & Lewandowsky, S. (2004). Verbalizing Facial Memory: Criterion Effects in Verbal Overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 739–755. doi: 10.1037/0278-7393.30.4.739.

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17,* 629 – 654. doi: 10.1002/acp.891

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological Science and Public Policy. *Perspectives on Psychological Science, 7,* 238 – 259. doi: 10.1177/1745691612439584

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior, 35*(5), 364–380. doi: 10.1007/s10979-010-9245-1.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*, 37–60. doi: 10.3758/BF03210740.

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. doi: 10.1177/0956797616655789

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law, 19,* 151 – 164. doi: 10.1037/a0030618

Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior, 43*(5), 436–454. doi: 10.1037/lhb0000343.

Lewandowsky, S. & Farrell, S. (2011). *Computational modeling in Cognition: Principles and Practice.* Sage Publications.

Lindsay, R. C. L. & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4,* 303 – 313. doi: 10.1007/bf01040622

RStudio Team (2015). *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.*

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. doi: 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. doi: 10.1177/0956797617698528

Smith, A. M., Yang, Y., & Wells, G. L. (2020). Distinguishing between investigator discriminability and eyewitness discriminability: a method for creating full receiver operating characteristic curves of lineup identification performance. *Perspectives on Psychological Science, 15,* 589 – 607. doi: 10.1177/1745691620902426

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

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior, 44*, 3-36. doi: 10.1037/lhb0000359

Wells, G. L., & Luus, E. (1990). Police lineups as experiments: Social methodology as a framework for properly-conducted lineups. *Personality and Social Psychology Bulletin, 16,* 106 – 117. doi: 10.1177/014167290161008

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 and Cognition, 4,* 313 – 317. doi: 10.1016/j.jarmac.2015.08.008

Wells, G. L., Small, M., Penrod, S. J., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22*, 603-647. doi: 10.1023/A:1025750605807

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 and Cognition, 4,* 324 – 328. doi: 10.1016/j.jarmac.2015.08.010

Wetmore, S.A., McAdoo, R.M., Gronlund, S.D. *&* Neuschatz, J. S. (2017). The impact of fillers on lineup performance. *Cognitive Research: Principles and Implications* **2,** 48. doi: 10.1186/s41235-017-0084-1

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition, 4,* 8 -14. doi: 10.1016/j.jarmac.2014.07.003

Wickham, H. (2017). Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://CRAN/R-project.org/package=tidyverse

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

Wixted, J. T., & Mickes, L. (2015a). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition, 4*(4), 318–323. doi: 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 and Cognition, 4*(4), 329–334. doi: 10.1016/j.jarmac.2015.08.007

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

Supplemental Materials

In the following document, we present additional analyses that we excluded from the main body for brevity. We present supplemental material for both Experiment 1 and for Experiment 2. In reference to Experiment 1, the primary focus of the supplemental materials is on (a) whether fair lineups produce better investigator discriminability than do biased lineups, and (b) how the confidence-accuracy characteristic curves vary as functions of lineup bias and memory strength. In reference to Experiment 2, we provide supplemental information that might help the more detailed-oriented reader to understand why constraining fillers to match the perpetrator on the same number of descriptors increases the magnitude of the fair-lineup advantage.

Experiment 1

Probit Regression Analyses

We used a series of probit regression analyses to test our *a priori* hypotheses that, (1) stronger memories increase perpetrator identifications and decrease mistaken identifications, (2) fair lineups decrease innocent-suspect picks more than they decrease perpetrator picks (two-way interaction) and, finally, that this effect is more pronounced under weak memory conditions than under strong memory conditions (3-way interaction). In our preregistration, we stated that we would use a logit link function (i.e., logistic regression), however, after collecting the data (but before looking at the data), we decided to use a probit link function instead. Although the two link functions almost always lead to the same conclusion (e.g., Agresti, 2002, 2007), we decided to use a probit regression because with appropriately specified models, the regression coefficients can be interpreted as d' scores because they give identical results (DeCarlo, 1998).

The Impact of Memory Strength on Perpetrator Identifications and Mistaken Identifications

We first tested the hypothesis that stronger memories increase perpetrator identifications and decrease mistaken identifications. To test this prediction, we created a dependent variable in which perpetrator identifications from the perpetrator present lineup were coded as 1s, both filler picks and innocent-suspect picks from the perpetrator absent lineup were coded as 1s, and all other outcomes were coded as 0s (i.e., perpetrator-present filler identifications, perpetrator-present rejections, and perpetrator-absent rejections). We regressed this dependent variable on memory strength, perpetrator presence, and the two-way interaction term. The interaction between memory strength and perpetrator presence was significant, $B = -0.54$, $SE = 0.14$, $Z = -3.93$, $p < .001$. We followed up this interaction by looking at the simple slopes of memory strength for perpetrator-present and perpetrator-absent lineups. The strong memory condition led to more perpetrator identifications than did the weak memory condition, $B = .19$, $SE = 0.10$, $Z = 2.00$, $p = .046$. The strong memory condition also led to fewer false positive identifications from the perpetrator-absent lineup than did the weak memory condition, $B = -0.34$, $SE = 0.10$, $Z = 3.55$, $p < .001$. Hence, our manipulation of memory strength (number of encoding trials) produced a strength-based mirror effect; the strong memory condition led to both more perpetrator identifications and fewer false-positive identifications than did the weaker memory condition (e.g., Glanzer & Adams, 1985; Smith et al., 2019; Smith, 2020).

The Impact of Lineup Similarity on Investigator Discriminability

To test the hypothesis that fair lineups would lead to better investigator discriminability (trade-off between perpetrator and innocent-suspect identifications) than would biased lineups and that this effect would be larger under weak memory conditions than under strong memory conditions, we used a hierarchical probit regression analysis. We entered all main effects on the first block, two-way interactions on the second block, and the three-way interaction on the third block. Because investigators do not perpetuate filler-identification errors, we treated perpetrator-absent filler picks as suspect rejections; hence, only perpetrator identifications and innocent-suspect identifications were coded as 1s, and all other lineup outcomes were coded as 0s. Not surprisingly, identifications of the suspect were more likely to occur when the suspect was guilty as opposed to innocent, $B = 0.76$, $SE = 0.075$, $Z = 10.18$, $p < .001$. The regression coefficient for perpetrator presence is equivalent to d' ($Z_{\text{Perpetrator IDs}} - Z_{\text{Innocent-Suspsect IDs}}$) and represents investigator discriminability for this experiment collapsed over all manipulations. In addition, suspect identifications were less likely to occur from a fair lineup than from a biased lineup, $B = -1.03$, $SE = 0.08$, $Z = 13.56$, $p < .001$. In other words, the fair lineup offered a more conservative test of the investigator's hypothesis that the suspect was the perpetrator.

Next, we examined all two-way interactions. Neither the interaction between memory strength and lineup bias, $B = -0.03$, $SE = 0.15$, $Z = -0.23$, $p = .82$, nor the interaction between perpetrator-presence and lineup bias, $B = 0.25$, $SE = 0.15$, $Z = 1.62$, $p = .10$, was significant. The lack of an interaction between perpetrator presence and filler similarity is noteworthy, as it suggests that fair fillers did not significantly improve the capacity of investigators to discriminate between perpetrators and innocent suspects (though the effect was "trending" in that direction). We did find a significant interaction between perpetrator-presence and memory strength, $B = -0.35$, $SE = 0.15$, $Z = -2.32$, $p = .02$. Because the three-way interaction was not significant, $B = 0.10$, $SE = 0.31$, $Z = 0.33$, $p = .74$, we followed up the significant two-way interaction between perpetrator presence and memory strength by examining the simple slopes of perpetrator-presence (d') at different levels of memory strength. Under strong memory conditions, perpetrator identifications were significantly more likley to occur than were innocent-suspect identifications, $B = 0.83$, $SE = 0.12$, $Z = 6.89$, $p < .001$. Likewise, perpetrator identifications were significantly more likley to occur than were innocent suspect identifications under weak memory conditions, though the different was relatively weaker $B = 0.48$, $SE = 0.12$, $Z = 3.86$, $p < .001$. The regression coefficients for these simple slopes are equivalent to the d' values for strong and weak memory conditions, respectively. Hence, these results suggest that the capacity of the investigator to discriminate between perpetrators and innocent suspects (d') was superior when eyewitnesses had strong memories compared to when eyewitnesses had weak memories.

Receiver Operating Characteristic Analyses

For simplicity, in the main body of our paper, we focused on the capacity of fair and biased lineups to sort only between guilty-suspect identifications and innocent-suspect identifications to the exclusion of other lineup outcomes (filler identifications and not-present responses). That has been the primary focus of the eyewitness identification field for more than 40 years, and for over 40 years, we have known that fair lineups better discriminate between guilty- and innocent-suspect identifications than do biased lineups (Lindsay & Wells, 1980). Hence the field is on good footing to debate theoretical accounts for why fair lineups better sort guilty-suspect identifications from innocent-suspect identifications than do biased lineups. In other work, we have called into question the myopic focus of eyewitness researchers on suspect identifications (e.g., Smith et al., 2020; Wells & Lindsay, 1980; Wells et al., 2015, Yang et al., 2019). Lineups not only have the capacity to “rule in” guilty persons, they also have the capacity to “rule out” innocent persons. Obviously, eyewitness scientists should be concerned both with the relative capacities of lineup procedures to “rule in” the guilty as well as their relative capacities to “rule out” the innocent. Indeed, we cannot make policy recommendations about which of two lineup procedures should be preferred unless we consider both their relative capacities to rule in the guilty and to rule out the innocent. We recently introduced a method for creating full ROC curves of eyewitness lineups for the purpose of comparing the total informational value of two eyewitness lineup procedures (Smith et al., 2020), and we use that tool here to compare fair and biased lineups.

Recall that participants in our first experiment who initially made an affirmative identification (i.e., of a suspect or a filler) were asked to provide a confidence statement about their identification. Participants who initially rejected the lineup were asked to select the lineup member who best matched their memory for the perpetrator, after which they were asked to indicate how confident they were that this individual was the perpetrator. In order to build our ROC curves, we used both initial suspect and filler picks along with final forced-choice suspect and filler picks.

First, we binned each respective identification decision (suspect picks, filler picks, forced-choice suspect picks, and forced-choice filler picks) into three levels of confidence: high (90% - 100%), medium (70% - 80%), and low (0% to 60%). Then, to create our ROC curves, we plotted high-confidence perpetrator identifications (on the Y-axis) against high-confidence innocent-suspect identifications (on the X-axis) in the ROC space. This was our first and most stringent operating point. Next, we plotted the proportions of perpetrator and innocent-suspect identifications made with at least medium confidence in the ROC space. Hence, this second-to-rightmost operating point includes all of the participants who made a suspect identification with high confidence plus all of the participants who made a suspect identification with medium confidence. The next operating point after that plotted the cumulative perpetrator and innocent-suspect identification rates collapsed across all levels of confidence.

After plotting all of the initial suspect identifications in the ROC space, we continued our ROC curve by plotting forced-choice suspect identifications in descending order of confidence levels. The logic for including the forced-choice suspect identifications next was that these eyewitnesses preferred the suspect over the other lineup members, but had too stringent a criterion to make an affirmative identification. Next, we plotted filler identifications in *ascending* levels of confidence and finally, we plotted forced-choice filler identifications in *ascending* levels of confidence. We made the decision to generate our ROC curves in this manner prior to examining the ROC curves, as our objective was to order the operating points from strongest to weakest evidence of guilt as one pans from left to right in the ROC space.

We started our ROC analyses by comparing strong and weak memory conditions collapsed over filler similarity and by comparing fair lineups and biased lineups collapsed over memory strength. Investigator discriminability was superior when eyewitnesses were given strong memories (AUC = .69) compared to when eyewitnesses were given weak memories (AUC = .59), $D = 3.28$, $p = .001$. Consistent with the results of our probit regression analyses, investigator discriminability did not significantly differ as a function of filler similarity (fair AUC = .67; biased AUC = .66), $D = 0.54$, $p = .59$.

We then examined what might be thought of as simple main effects. Specifically, we examined the strong and weak memory conditions separately for fair and biased lineups and we also compared fair lineups and biased lineups at both strong and weak levels of memory strength. These four contrasts are presented in the four quadrants of Figure S1. For eyewitnesses who were shown a fair lineup, strong memory conditions led to a non-significant increase in investigator discriminability (AUC = .70) relative to weak memory conditions (AUC = .63), $D = 1.61$, $p = .11$. For eyewitnesses who were shown a biased lineup, strong memory conditions significantly increased investigator discriminability (AUC = .70) relative to weak memory conditions (AUC = .61), $D = 2.25$, $p = .02$. The impact of memory strength on investigator discriminability is straightforward: Eyewitnesses with better memories provide the investigator with better evidence, which increases the capacity of the investigator to discriminate between perpetrators and innocent suspects.

Although stronger memories increased investigator discriminability relative to weaker memories, fair fillers did not increase investigator discriminability relative to biased fillers at either weak (AUC = .63 vs. AUC = .61), $D = 0.57$, $p = .57$, or strong (AUC = .70 vs. AUC = .70) levels of memory strength, $D = 0.08$, $p = .94$. Although this finding is inconsistent with our pre-registered hypotheses, it is consistent with a very recent reanalysis of Colloff et al.'s (2016) data that we published since preregistering this experiment; indeed, even for the original dataset, the full ROC curve demonstrates that fair lineups did not lead to a better tradeoff than did biased lineups (Smith et al., 2020).

The finding that fair lineups did not increase the area under the full ROC curve relative to biased lineups is a challenge to any theory or hypothesis that posits a fair-lineup advantage. Indeed, if fair lineups are truly better than biased lineups, then they should both be better able to "rule in" guilty suspects and to "rule out" innocent suspects. In the far-left region of the ROC space, fair lineups dominate biased lineups, suggesting increased capacity to "rule in" the guilty—the finding we discussed in our main manuscript, which has been the primary focus of the eyewitness field over the last 40 years. But as we move further to the right in the ROC space, biased lineups begin to dominate fair lineups, suggesting that biased lineups may be more effective at "ruling out" innocent suspects. There are a number of possible explanations for this pattern of results, but we leave further discussion of this problem to future research.

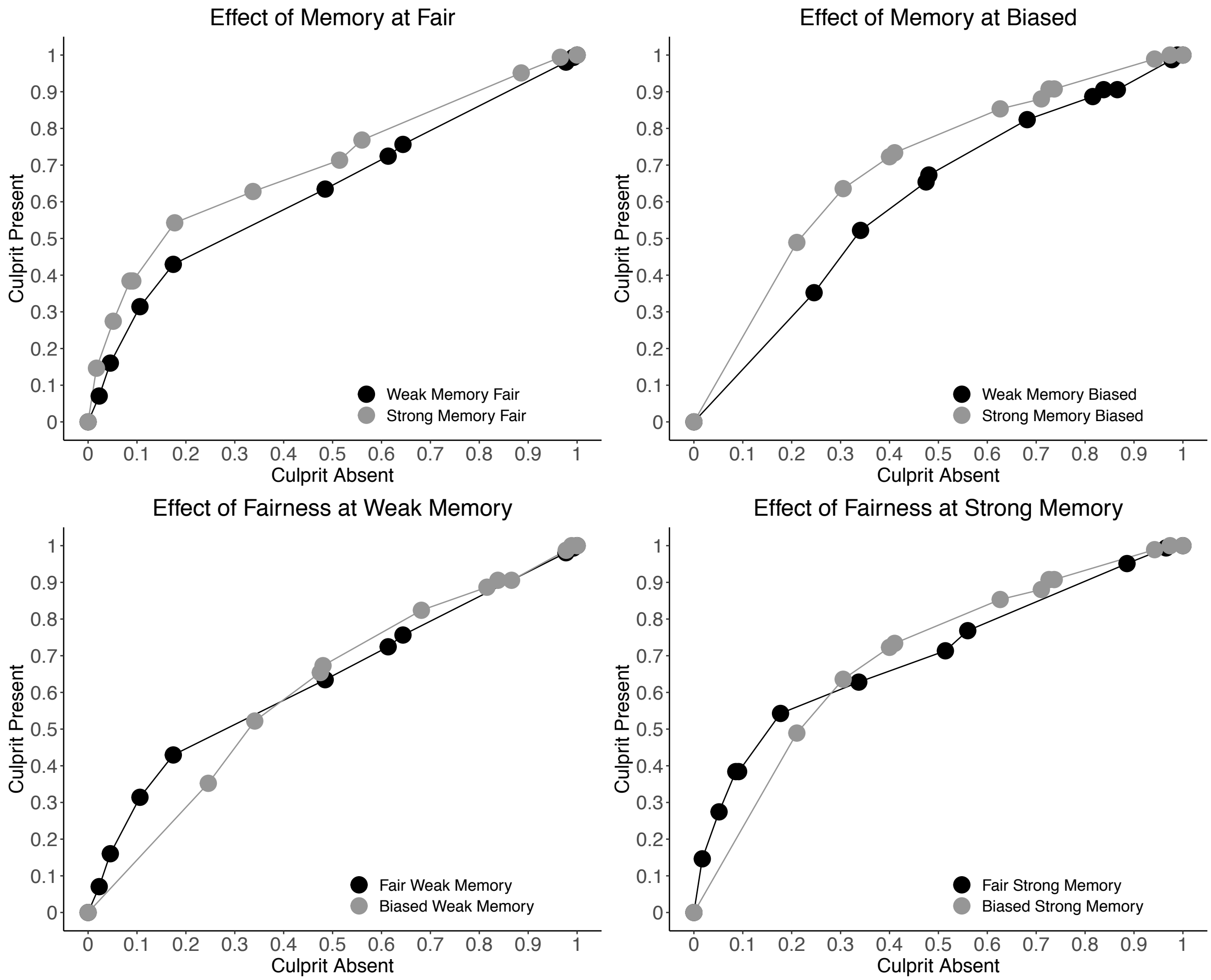


Figure S1. Receiver Operating Characteristic (ROC) analyses comparing weak-memory fair lineups with strong-memory fair lineups, weak-memory biased lineups with strong-memory biased lineups, fair weak-memory lineups with biased weak-memory lineups, and fair strong-memory lineups with biased strong-memory lineups.

Confidence-Accuracy Characteristic Curves

Finally, we used confidence-accuracy characteristic (CAC) curves to assess how well confidence could distinguish between accurate (perpetrator) identifications and inaccurate (innocent suspect) identifications (Mickes, 2015). When determining which procedure is superior, experimentalists tend to focus on discriminability, because it provides an estimate of performance that is independent of both decision criterion and the perpetrator-present base rate (e.g., Agresti, 2007). But, when it comes to an individual case, the criminal justice system uses eyewitness confidence to estimate the reliability of the eyewitness (e.g., Wells et al., 2015; Wixted & Wells, 2017). Hence, the forthcoming CAC analyses compliment (rather than compete with) the discriminability analyses that we have focused on to this point.

The CAC curves are presented in Figure S2 and the counts associated with each confidence bin are presented in Table S1. Because the counts in each bin are relatively small and because the confidence-accuracy relationship is not the primary focus of this work, we focus only on descriptive trends. The biggest takeaway from these curves is that suspect identifications from fair lineups were more reliable than suspect identifications from biased lineups across all confidence bins. Conversely, memory strength appears to have had less impact on suspect-identification accuracy; however, at high-confidence, suspect-identification accuracy is noticeably lower for weak compared to strong memory conditions.

There is one other noteworthy observation from these CAC curves. Across all conditions, high-confidence suspect identifications were not associated with high levels of accuracy (Giacona et al., 2021; Sauer et al., 2019; Smith et al., 2021; cf. Wixted & Wells, 2017). Even for the strong-memory fair lineup condition, the proportion of suspects identified with high confidence that were actually guilty was only about 90%. For the weak-memory fair lineup condition, high-confidence suspect identification accuracy dropped to 81%. Moreover, confidence was extremely unreliable for biased lineups: high-confidence suspect identifications from the strong-memory biased lineup condition were correct only 69% of the time and high-confidence suspect identifications from the weak-memory biased lineup condition were correct only 56% of the time.

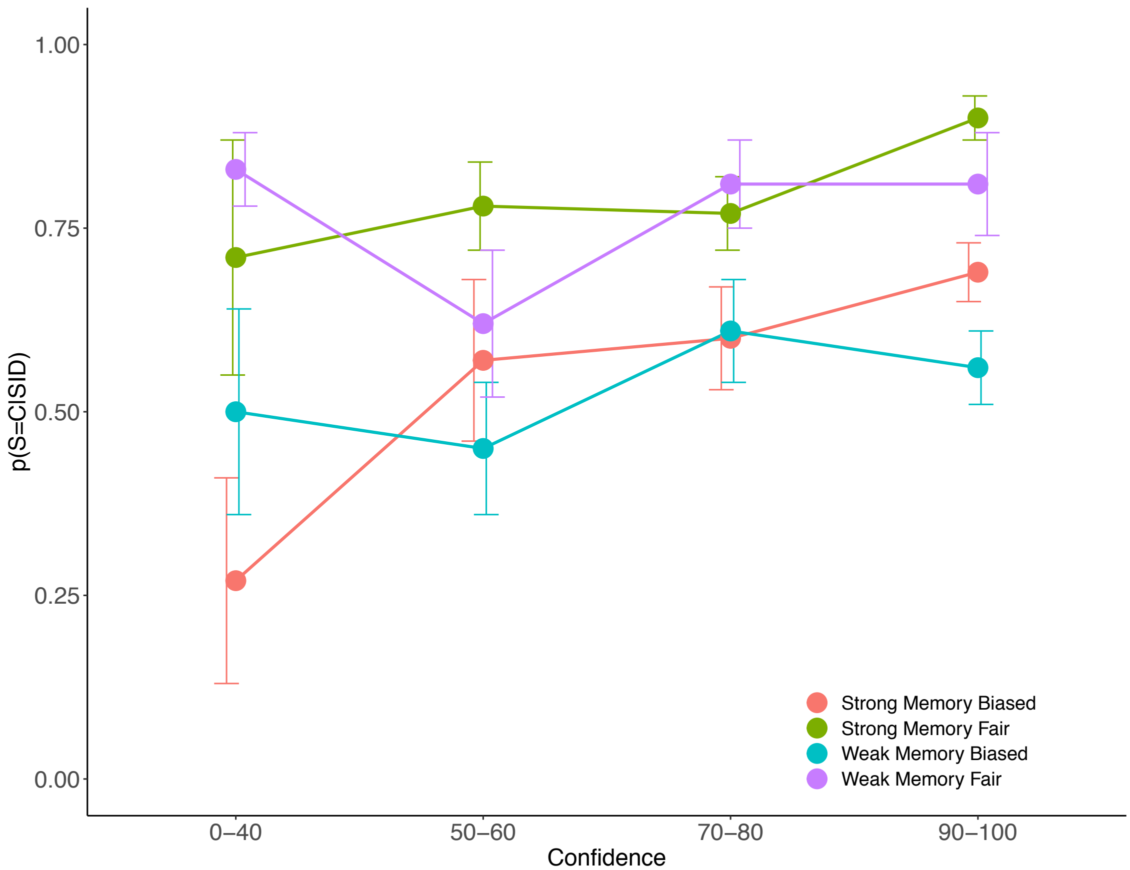


Figure S2. Confidence-accuracy-characteristic curves as a function of filler similarity and memory strength.

Table S1

Counts for Each Bin in the Confidence-Accuracy Characteristic Curves Presented in Figure S2

Procedure	TP/TA	0 - 40	50 – 60	70 – 80	90 - 100
Strong Biased	TP	3	13	27	90
	TA	8	10	18	40
Strong Fair	TP	4	14	21	24
	TA	10	24	37	16
Weak Biased	TP	7	14	27	56
	TA	7	17	17	44
Weak Fair	TP	16	8	14	11
	TA	20	29	20	15

Note. For the fair perpetrator-absent lineups, the total false-positive rate is provided. The innocent-suspect identification rate can be found by dividing these values by six (the number of lineup members).

Experiment 2

Why Does the Magnitude of the Fair-Lineup Advantage Increase When Fillers are Constrained to Match the Perpetrator on the Same Descriptors?

Statistical independence refers to (among other things) situations in which the variability within a group (or cluster) is the same as the variability across groups (or clusters). If there is less variability within groups than across groups, we reject the assumption of statistical independence and assume statistical dependence. In the context of eyewitness lineup procedures, this boils down to the question of whether the probability of being identified is more similar for members *within* lineups than for *members* between lineups. If we can assume that the probability of being identified for members *within* lineups is no more similar than the probability of being identified for members *between* lineups, then we can assume statistical independence. If not, then we must reject the assumption of independence and account for statistical dependencies in our models.

Whether we assume statistical independence or statistical dependence is of critical import because all past instantiations of computational models that have failed to find evidence that differential filler-siphoning can explain the fair-lineup advantage have assumed statistical independence (e.g., Colloff et al., 2016; Wixted & Mickes, 2015). Conversely, all past instantiations of computational models that have relaxed the assumption of independence have found that differential filler-siphoning can account for the fair-lineup advantage (Smith et al., 2017; Wetmore et al., 2017). And in fact, there is growing consensus in the field that the assumption of statistical independence is likely not viable (e.g., Akan et al., 2020; Smith et al., 2017; Wixted et al., 2018). As we discuss in the main body of this paper, there are several reasons why statistical independence is not viable, but one simple explanation has to do with the manner in which investigators construct fair lineups. When constructing a fair lineup, an investigator might select innocent persons, in part, because they possess each of the features that the witness gave in her description of the perpetrator. When an eyewitness provides a rich and detailed description, innocent members might match the perpetrator on a number of features. When an eyewitness provides a more impoverished description, innocent members might match the perpetrator on fewer features. Hence, we should expect less variability in signal strengths *within* lineups than between lineups. This is how we constructed lineups in our instantiation of WITNESS and under this condition, WITNESS demonstrated differential filler-siphoning and a fair lineup advantage.

Differential filler-siphoning refers to the phenomenon that fair lineup fillers siphon more identifications away from innocent suspects than they do from guilty suspects. The discrepancy between how many guilty versus innocent-suspect identifications are siphoned away by fillers is increased under the assumption of dependence compared to under the assumption of independence. Whether one assumes independence or not, fair fillers always siphon 5 in every 6 affirmative identifications away from the innocent suspect. But models that assume independence assume that fillers siphon more identifications away from the perpetrator than do models that assume dependence. In other words, models that assume independence assume that the magnitude of the differential filler-siphoning phenomenon, and thus the size of the fair-lineup advantage, is smaller than do models that allow for dependence.

Table S2 helps to conceptualize why models that assume independence predict more siphoning from the perpetrator than do models that assume dependence. Note that this is a strictly conceptual demonstration and for sake of clarity, the demonstration exaggerates the difference between assumptions of independence and dependence. For simplicity, we assume four-person lineups with one perpetrator and three fillers. S2 displays three hypothetical lineups for cases involving different perpetrators. Under the assumption of independence, the high-similarity, moderate-similarity, and low-similarity fillers are equally distributed across the three cases. Under the assumption of dependence, all fillers in the first lineup are high in similarity, all fillers in the second lineup are moderate in similarity, and all fillers in the third lineup are low in similarity to the perpetrator. Because fillers who are high in similarity to the perpetrator are more likely to draw identifications away from the perpetrator, the independent-assumption lineups in Table S2 assume that there is always someone who could strongly compete for identifications with the perpetrator. Conversely, the dependent-assumption lineups in Table S2 assume that all the strong competitors are present in the first lineup and that the fillers in the second and third lineups would not be able to compete as strongly for identifications with the perpetrator. Hence, one way to conceive why the assumption of independence predicts more siphoning of perpetrator identifications than does the assumption of dependence is because high-similarity fillers who can strongly compete for identifications with the perpetrator are equally likley to occur across lineups, creating more opportunities to siphon identifications from perpetrators. Conversely, under the assumption of dependence, those strong competitors tend to cluster in the same lineups, creating fewer opportunities to siphon from perpetrators overall.

Table S2

Conceptual Demonstration of Why Models that Assume Independence of Signal Strength Predict More Siphoning from Perpetrators Than Do Models that Assume Dependence of Signal Strength

	Independence				Dependence			
	Perp	Filler 1	Filler 2	Filler 3	Perp	Filler 1	Filler 2	Filler 3
Case 1	P	H	M	L	P	H	H	H
Case 2	P	H	M	L	P	M	M	M
Case 3	P	H	M	L	P	L	L	L

Note. P = perpetrator; H = High-Similarity Filler; M = Moderate-Similarity Filler; L = Low-Similarity Filler

Again, we stress that the example in Table S2 is hyperbolic as we have intentionally created a situation where, under the assumption of independence, fillers of different similarity values are perfectly spread across lineups, and under the assumption of dependence, fillers of different similarity values are perfectly sorted into lineups. This is an oversimplification, but the intention is to help the reader to understand why the independence assumption predicts more siphoning from the perpetrator, and why, under the assumption of dependence, models tend to predict that the magnitude of differential filler-siphoning, and thus the fair-lineup advantage, will be larger. Finally, because innocent lineup members are selected based on matching the perpetrator on a common set of features, because encoding strength does vary from one witness to the next, and because decision criteria vary from one witness to the next, the independence assumption is not viable for lineup models. Clearly, resting the assumption of independence has very important implications for the conclusions we draw.

References

Agresti (2002). *Categorical Data Analysis* (2nd ed.). Hoboken, New Jersey: Wiley.

Agresti (2007). *An Introduction to Categorical Data Analysis* (2nd ed.). Hoboken, New Jersey: Wiley.

Akan, M., Robinson, M. M., Mickes, L., Wixted, J. T., & Benjamin, A. S. (2021). The effect of lineup size on eyewitness identification. *Journal of experimental psychology. Applied*, *27*, 369–392. doi: 10.1037/xap0000340.

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. doi: 10.1177/0956797616655789

DeCarlo, L. T. (1998) Signal detection theory and generalized linear models. *Psychological Methods, 3,* 186 – 205. doi: 10.1037/1082-989X.3.2.186

Giacona, A. M., Lampinen, J. M., & Anastasi, J. S. (2021). Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law and Human Behavior, 45*(3), 256–270. doi: 10.1037/lhb0000381.

Glanzer, M., Adams, J.K. (1985). The mirror effect in recognition memory. *Memory & Cognition 13***,** 8–20. doi: 10.3758/BF03198438.

Lindsay, R. C. L. & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4,* 303 – 313. doi: 10.1007/bf

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4,* 93 – 102. doi: 10.1016/j.jarmac.2015.01.003

Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence todiagnose the accuracy of individual identification decisions. *Psychology, Public Policy, and Law, 25*, 147 – 165. doi: 10.1037/law0000203

Smith, A. M. (2020). Why do mistaken identifications increase when either witnessing or testing conditions get worse? *Journal of Applied Research in Memory and Cognition, 9,* 495 – 507. doi: 10.1016/j.jarmac.2020.08.002

Smith, A. M., Smalarz, L., Ditchfield, R., & Ayala, N. T. (2021). Evaluating the claim that high confidence implies high accuracy in eyewitness identification. *Psychology, Public Policy, and Law.* Advance online publication. doi: 10.1037/law0000324.

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. doi: 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. doi: 10.1177/0956797617698528

Smith, A. M., Wilford, M., Quigley-McBride, A., & Wells, G. L. (2019). Mistaken eyewitness identification rates increase when either witnessing or testing conditions get worse. *Law and Human Behavior, 43,* 358 – 368. doi: 10.1037/lhb0000334

Smith, A. M., Yang, Y., & Wells, G. L. (2020). Distinguishing between investigator discriminability and eyewitness discriminability: a method for creating full receiver operating characteristic curves of lineup identification performance. *Perspectives on Psychological Science, 15,* 589 – 607. doi: 10.1177/1745691620902426

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88*, 776–784. 10.1037/0033- 2909.88.3.776

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior, 39*, 99–122. doi: 10.1037/lhb0000125.

Wetmore, S.A., McAdoo, R.M., Gronlund, S.D. *&* Neuschatz, J. S. (2017). The impact of fillers on lineup performance. *Cognitive Research: Principles and Implications* **2,** 48. doi: 10.1186/s41235-017-0084-1

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

Wixted, J. T. & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18,* 10 – 65. doi: 10.1177/152910061668696

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

Yang, Y., Smalarz, L., Moody, S. A., Cabell, J. J., & Copp, C. J., (2019). An expected cost model of eyewitness identification. *Law and Human Behavior, 43,* 205 – 219. doi: 10.1037/lhb0000331.